

EXHIBIT 16

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF ILLINOIS
EASTERN DIVISION

ANDREW CORZO, SIA HENRY, MICHAEL
MAERLANDER, ALEXANDER LEO-
GUERRA, BRANDON PIYEVSKY,
BENJAMIN SHUMATE, BRITTANY
TATIANA WEAVER, and CAMERON
WILLIAMS, individually and on behalf of all
others similarly situated,

Plaintiffs,

v.

BROWN UNIVERSITY, CALIFORNIA
INSTITUTE OF TECHNOLOGY,
UNIVERSITY OF CHICAGO, THE TRUSTEES
OF COLUMBIA UNIVERSITY IN THE CITY
OF NEW YORK, CORNELL UNIVERSITY,
TRUSTEES OF DARTMOUTH COLLEGE,
DUKE UNIVERSITY, EMORY UNIVERSITY,
GEORGETOWN UNIVERSITY,
MASSACHUSETTS INSTITUTE OF
TECHNOLOGY, NORTHWESTERN
UNIVERSITY, UNIVERSITY OF NOTRE
DAME DU LAC, THE TRUSTEES OF THE
UNIVERSITY OF PENNSYLVANIA,
WILLIAM MARSH RICE UNIVERSITY,
VANDERBILT UNIVERSITY, and YALE
UNIVERSITY,

Defendants.

Case No. 1:22-cv-00125

Class Action

**REBUTTAL EXPERT REPORT OF
HAL J. SINGER, PH.D.**

October 7, 2024

This report cites and quotes material that Defendants designated as
Confidential or AEO under the Second Amended Confidentiality Order (ECF No. 608)

TABLE OF CONTENTS

	<u>Page</u>
INTRODUCTION AND SUMMARY OF CONCLUSIONS	1
BACKGROUND.....	11
I. COLLECTIVE MARKET POWER.....	12
A. Dr. Hill’s Critique of My Hypothetical Monopolist Test Is Invalidated by My Regression Results	13
B. Defendants’ Experts Claim That Defining the Relevant Market As a Subset of Universities Ranked Highly by USNWR Is Not Appropriate	14
1. Dr. Hill and Dr. Long Incorrectly Claim USNWR Rankings Are the Basis for My Market Definition.....	14
2. Dr. Hill Incorrectly Throws out Any Consideration of the USNWR Rankings for Market Definition	14
C. Academic Literature Does Not Undermine My Reliance on USNWR Rankings	17
1. The Hu Paper Is Not Inconsistent with My Use of the USNWR Rankings.....	17
2. The Rankings from the Avery et al. Paper Have Not Displaced the USNWR Rankings	18
3. Dr. Hill Claims That Applying Hu’s Revealed Preference Method Using Parchment.com Data Are Inconsistent with the Market Definition.....	18
4. Dr. Long’s Claim That Some Students Applying to a Wide Range of Schools Implies a Broader Relevant Market Here Is Wrong	19
D. Defendants’ Experts Are Wrong That Public Schools Belong In The Relevant Market.....	20
1. Dr. Hill Incorrectly Cites the Application Behavior of Certain Named Plaintiffs as Evidence That Public Universities Belong in the Relevant Market	20
2. Dr. Hill Incorrectly Asserts That Defendants’ Own Documents Show Public Universities Belong in the Relevant Market.....	21
II. THE CHALLENGED CONDUCT.....	23
A. Contrary to Defendants’ Experts’ Arguments, There Is Substantial Evidence That All Defendants Imposed the Consensus Methodology.....	25
B. Defendants’ Experts Misleadingly Argue That Defendants’ Actions Are Consistent with The Rest of The Higher Education Industry	34
III. COMMON IMPACT	40

A.	Classwide Evidence Is Consistent with the Challenged Conduct Causing an Artificial Inflation in Effective Institutional Prices for All Defendants	41
1.	Qualitative Evidence Supports the Claim that the Challenged Conduct Would Have Resulted in Artificially Inflated Effective Institutional Prices	41
2.	Defendants' Experts Critiques of My Overcharge Model Are Without Merit.....	48
B.	Classwide Evidence Demonstrates that the Overcharge Would Have Impacted All or Nearly All Class Members	130
1.	Dr. Stiroh's Principle-by-Principle Approach to Assessing Classwide Impact Is Fundamentally Flawed	131
2.	Econometric Evidence Confirms That the Generalized Effective Institutional Price Overcharges Were Paid by All or Nearly All Class Members.....	137
C.	Defendants' Experts Do Not Identify Any Procompetitive Benefits to Justify the Existence of the 568 Group	147
IV.	AGGREGATE DAMAGES.....	151
	APPENDIX 1: MATERIALS RELIED UPON.....	153
	APPENDIX 2: DR. HILL'S QUALITATIVE EVIDENCE REGARDING DEFENDANTS' DIFFERING CALCULATIONS APPEARS TAKEN OUT OF CONTEXT AND IS INCONSISTENT WITH HIS CONCLUSIONS	168
A.	Qualitative Evidence Does Not Support That Defendants Calculated Family Contributions Differently	168
B.	Dr. Hill's Qualitative Evidence Regarding Net Price Calculations (Section 5.2.4).....	176
	APPENDIX 3: DEFENDANTS' EXPERTS' PREPACKAGED ANALYSES ARE NOT PERSUASIVE	179
A.	Irrelevant Prepackaged Analyses	182
1.	Dr. Hill Presents Isolated Examples Under the Guise of Analysis.....	182
2.	Dr. Hill Uses Faulty Benchmarks in His Prepackaged Analyses.....	182
3.	Dr. Hill Repeatedly Misconstrues the Relevant Counterfactual.....	183
B.	Relevant Prepackaged Analyses	184
1.	Dr. Hill's Assertion That Defendants Increased Financial Aid at the Same Pace as Peer Schools Is Misleading	184
2.	Dr. Stiroh's Review of Cross-Admitted Students Is Without Merit.....	190
3.	Dr. Hill's Cross-Admit EFC Analysis Does Not Address The Challenged Conduct.....	191

4.	Dr. Hill Repeats His Inadequate Cross-Admit Analysis With Net Price.....	193
C.	Defendants’ Experts’ Claims That The Consensus Methodology Did Not Standardize Financial Aid Are Without Merit.....	194
1.	Dr. Hill’s Argument That Defendants Do Not Have Matching Net Prices Is Irrelevant and Misleading.....	194
2.	Dr. Hill’s Claim That Defendants Did Not Change Their Financial Aid When Entering or Exiting the 568 Group Is Incorrect	195

APPENDIX 4: STATISTICAL ANALYSES OF THE REMAINING DEFENDANTS’

	WEALTH FAVORITISM IN ADMISSIONS	196
A.	Cornell.....	196
B.	Georgetown.....	201
C.	MIT	205
D.	Notre Dame.....	209
E.	Penn.....	213

INTRODUCTION AND SUMMARY OF CONCLUSIONS

1. Counsel for Plaintiffs have asked me to review the expert reports of Dr. Nicholas Hill,¹ Dr. Lauren Stiroh,² Dr. Bridget Terry Long,³ and Dr. David L. Yermack⁴ (collectively, “Defendants’ Experts”) issued in response to my June 10, 2024 Expert Report in this matter (“Initial Report” or “Singer Report”).⁵ The vast majority of their criticisms lack any merit. As to the handful of critiques that have some merit, I have appropriately incorporated them into my analyses. Doing so does not undermine any of my opinions, though my conclusion regarding the amount of aggregate damages to the Class has been revised.⁶ Several of Defendants’ Experts offer certain affirmative opinions that are not responsive to any of my analyses. Each of these opinions is flawed, as I show below.

2. In my Initial Report, I reached four major opinions. Below, I summarize each of these opinions. I also partially summarize Defendants’ Experts’ critiques, my concessions to these opinions, and my responses.

- (1) **Collective Market Power**. In my Initial Report, I found that Defendants held collective market power in the relevant antitrust market for academic services that Elite Private Universities provide to undergraduate students during the Class Period.⁷ I define the Elite Private Universities as private universities ranked in the average top 25 U.S. News and World Report rankings over the period 2003-2022. This opinion is supported by both direct and indirect evidence.

I provided direct quantitative evidence of Defendants’ collective market power by showing that Defendants were able to artificially inflate Effective

1. Expert Report of Nicholas Hill, Ph.D. (Aug. 7, 2024) [hereafter Hill Report].

2. Expert Report of Lauren J. Stiroh, Ph.D. (Aug. 7, 2024) [hereafter Stiroh Report].

3. Expert Report of Bridget Terry Long, Ph.D. (Aug. 7, 2024) [hereafter Long Report].

4. Expert Report of David L. Yermack (Aug. 7, 2024) [hereafter Yermack Report].

5. Errata II Expert Report of Hal J. Singer, Ph.D. (Jun. 10, 2024). I use the terminology “Initial Report” when referring to this report within the body text, and otherwise refer to it as “Singer Report” in footnote citations. I submitted my initial expert report on May 14, 2024. On May 28, 2024, I submitted an errata to my Initial Report that (1) incorporated new data produced by Defendants after my initial submission, and (2) corrected “stale” results that had been discovered while preparing my initial workpapers production. On June 3, 2024, Defendants transmitted a letter identifying a purported “sort-order” issue in my workpapers. On June 4, 2024, Defendants produced an updated student UID crosswalk (UID stands for “unique student identifier”), which update replaced apparently incorrect versions of student UIDs with updated versions. I therefore submitted a second errata incorporating the changes resulting from the June 3 and June 4 communications, in addition to the two changes that had been previously reflected in the first errata.

6. I use the same defined terms here as I had used in my Initial Report.

7. Singer Report §I.

Institutional Prices above levels that would have prevailed absent the Challenged Conduct.⁸ I then provided indirect evidence by demonstrating that Elite Private Universities constitute a relevant market, showing that Defendants collectively held high market shares in said market, and showing that there are high barriers to entry in the relevant market.⁹ To show that Elite Private Universities constitute a relevant antitrust market, I first conducted a hypothetical monopolist test (“HMT”), which shows that Defendants were able to artificially inflate Effective Institutional Prices by an economically significant degree and over more than a decade.¹⁰ I then described practical indicia known as the *Brown Shoe* factors which form the basis of my market definition.¹¹ The *Brown Shoe* factors that I provide include: (i) qualitative evidence showing that Defendants distinguish themselves and their peers as elite, private universities distinct from other post-secondary education providers in the United States,¹² (ii) a peer analysis that shows that institutions in the relevant market perceive themselves as distinct from institutions outside of the relevant market,¹³ (iii) an analysis showing that students attending universities in the relevant market tend to travel further distances than undergraduates attending universities not in the relevant market,¹⁴ and (iv) a revealed-preference analysis of pairwise comparisons of undergraduate admissions showing that undergraduates that are admitted to an institution in the relevant market consider and prefer other institutions in the relevant market more than institutions not in the relevant market in their enrollment decision.¹⁵ After having defined the relevant market of Elite Private University Services, I then showed that Defendants enjoyed sufficiently large shares of the relevant market to support the structural presumption of market power.¹⁶ Further, I discussed the barriers to entry that characterize the market for Elite Private University services.¹⁷

Defendants’ Experts take issue with my use of USNWR as a starting point for my market definition, my HMT, and my calculation of Defendants’ collective market shares. As I show below, Defendants’ Experts’ critiques are without

8. *Id.* §I.A.

9. *Id.* §I.B.

10. *Id.* §I.B.1.b. Economists consider both economic significance and statistical significance when interpreting results. Economic significance pertains to the overall size and sign of the coefficient. Statistical significance means that it is highly unlikely that the true value of the variable is equal to zero given such a large coefficient. JEFFREY WOOLDRIDGE, APPLIED ECONOMETRICS, A MODERN APPROACH (5th ed. 2012) [hereafter WOOLDRIDGE] at 135-36 (“Because we have emphasized statistical significance throughout this section, now is a good time to remember that we should pay attention to the magnitude of the coefficient estimates in addition to the size of the t statistics. The statistical significance of a variable x_j is determined entirely by the size of $t_{\hat{\beta}_j}$, whereas the economic significance or practical significance of a variable is related to the size (and sign) of $\hat{\beta}_j$. [] Too much focus on statistical significance can lead to the false conclusion that a variable is ‘important’ for explaining y even though its estimated effect is modest.”).

11. Singer Report §I.B.1.c.

12. *Id.* §I.B.1.a.

13. *Id.* §I.B.1.d.

14. *Id.* §I.B.1.e.

15. *Id.* §I.B.1.f.

16. *Id.* §I.B.2.

17. *Id.* §I.B.3.

merit. Moreover, Defendants' Experts do not address the evidence I provide establishing the Elite Private University Market using qualitative evidence under the *Brown Shoe* test and documenting the significant barriers to entry in the Elite Private University market. Dr. Hill asserts that USNWR is not a proper foundation for defining a market because it is purportedly subject to manipulation by universities, because it does not consider all characteristics relevant to students when deciding where to attend, and because USNWR shifts the characteristic weights that it uses to form rankings.¹⁸ Dr. Long similarly contends that university rankings more generally do not reflect all a student considers when choosing where to attend.¹⁹ Dr. Long also claims my ranking cutoff does not align with the industry or academic consensus.²⁰ Dr. Hill asserts that my HMT is flawed and unreliable because he finds no evidence of an artificial overcharge in Effective Institutional Prices attributable to the Challenged Conduct.²¹ Dr. Hill also takes issue with my calculation of Defendants' collective market share, claiming that I should have considered Defendants' participation in and out of the Challenged Conduct when computing Defendants' collective market share.²²

Defendants' Experts' critiques of USNWR rankings are without merit largely because they focus on the objectivity of the rankings rather than their centrality to students' and universities' assessment of school quality. The latter is what determines demand substitution, thereby informing market definition. Dr. Hill's critique of my HMT is predicated on his erroneous assertion that my Effective Institutional Price overcharge regressions show a non-statistically significant or non-positive overcharge. He can only obtain a non-statistically significant overcharge after making numerous improper data-processing and erroneous modeling adjustments, which I explain in more detail below. Dr. Hill's argument that I improperly calculated Defendants' collective market share is also wrong; even if I were to calculate collective market shares using participation shares, Defendants still maintained a collective market share above 50 percent throughout the Class Period, a share which remains highly indicative of market power.

For the foregoing reasons, I find that classwide evidence is capable of proving that Defendants held market power in the Elite Private University market.

- (2) **Violation.** In my Initial Report, I showed that, when viewed through an economic lens, qualitative evidence in this matter is consistent with the Challenged Conduct reflecting collusive action and inconsistent with unilateral competitive conduct. I explained that the Challenged Conduct is comprised of an alleged Overarching Agreement comprised of consensus on six elements,

18. Hill Report ¶152.

19. Long Report §VI.

20. *Id.* §VI.

21. Hill Report ¶150.

22. *Id.* ¶151, ¶169.

including six core principles of awarding institutional aid, the primacy of awarding need based aid, the use of the Institutional Methodology (“IM”) as the basis for developing and using the Consensus Methodology (“CM”), the development and use of the CM itself, a manual for exercising Professional Judgment that was to be applied to exceptional circumstances, and the sharing of competitively sensitive information (“CSI”).²³ Of these six elements, Defendants’ Experts only contested whether the record evidence is consistent with agreement on the CM. Defendants’ Experts did not contest my economic valuation of the qualitative evidence pertaining to any of the other elements of the Challenged Conduct, including the exchange of competitively sensitive information on prices, the organizational elements that effectuated supracompetitive pricing, and inducements to support collusion, such as through regular meetings and surveys.²⁴ Each of these forms of evidence is consistent with the criteria that economists recognize as indicative of anticompetitive conduct.²⁵ I provided quantitative evidence via my overcharge regressions consistent with *all* Defendants having engaged in the alleged overarching conspiracy, and inconsistent with competition.²⁶ I also provided quantitative evidence in the form of Defendant-specific in-sample prediction consistent with each Defendant’s participation in the Challenged Conduct and inconsistent with each Defendant acting unilaterally.²⁷ I offered no opinion regarding whether the Challenged Conduct violated antitrust law.

As stated, Defendants’ Experts assert that Defendants did not uniformly implement the CM for determining EFCs.²⁸ Defendants’ Experts claim that the five other elements of the Challenged Conduct besides implementing the CM were merely industry standard, and that Defendants therefore would have engaged in the same activities absent the Overarching Agreement.²⁹

Defendants’ Experts’ argument that Defendants had not uniformly implemented the CM is irrelevant because, as an economic matter, for the alleged conspiracy to impose anticompetitive effects, including a common impact across the Class, there is no requirement that each aspect of each institution’s behaviors match exactly. As with all cartels, it is not necessary that there be perfect coordination across all aspects of a conspiracy for the cartel to impose anticompetitive effects or a common impact.³⁰ Defendants’ experts’ argument is also inconsistent with my empirical findings presented in this rebuttal, which show that Defendants’

23. Singer Report ¶148.

24. *Id.* §II.D.2.

25. *Id.* §II.D.1.

26. *Id.* §II.B.

27. *Id.* §II.C.

28. *See* Part II.A.

29. *See* Part II.B.

30. *Price Fixing, Bid Rigging, and Market Allocation Schemes: What They are and What to Look For*, U.S. DEPARTMENT OF JUSTICE (revised Feb. 2021), <https://www.justice.gov/d9/pages/attachments/2016/01/05/211578.pdf> at 2 (“It is not necessary that the competitors agree to charge exactly the same price, or that every competitor in a given industry join the conspiracy.”).

participation in the Challenged Conduct resulted in a convergence of EFCs and Effective Institutional Prices for Class Members. Defendants' Experts' claims that Defendants would have acted in the same way in a but-for world are inconsistent with economic theory, the economic rationale behind the formation of the 568 Group in the first place and why it had required an antitrust exemption, and other evidence I discuss in my Initial Report and below.

With regards to the quantitative evidence presented in my Initial Report consistent with *all* Defendants having engaged in the Challenged Conduct and with *each* Defendant's participation in the Challenged Conduct and inconsistent with each Defendant's competing against each other in an unfettered manner, Defendants' Experts assert that my findings are invalid. They base their opinions on a materially and improperly altered version of my statistical modeling. When so altered, Defendants' Experts' versions no longer yield a positive and statistically significant overcharge for the alleged Cartel as a whole or for each of the Defendant's individually. Defendants' Experts' argument is flawed because, as I explain in the common-impact summary below, many of their data-processing and modeling adjustments are unreliable and incorrect. Once I correct their erroneous adjustments, while accepting certain corrections, I once again obtain evidence of an economically and statistically significant overcharge in Effective Institutional Prices, and my model continues to produce a result consistent with the alleged Conspiracy and each Defendant's participation in it. In addition to my statistical modeling, my report points to qualitative analyses and economic theory that bolster my statistical findings in both respects.

For the foregoing reasons, I continue to find that the qualitative and quantitative evidence is consistent with Defendants, and each of them, acting in a coordinated manner and inconsistent with Defendants acting in their unilateral best interests.

- (3) **Common Impact.** In my Initial Report, I employed a standard, two-step approach to demonstrating common impact, both of which steps are common to the Class as a whole.

Step 1: In the first step of my two-step approach, I estimated a regression model that found that the Challenged Conduct caused Effective Institutional Prices at Defendant universities to increase above the levels that would have prevailed in its absence (that is, "but-for" the Challenged Conduct).³¹ I also provided qualitative evidence consistent with the Challenged Conduct resulting in a generalized artificial increase in Effective Institutional Prices.³²

31. *Id.* §III.A.2.

32. *Id.* §III.A.1.

Dr. Stiroh asserts that there is no causal nexus between the Challenged Conduct and a generalized artificial inflation in Effective Institutional Prices.³³ Her argument ignores the economics literature on cartels, which shows that when competitors consistently meet to discuss methods for charging Effective Institutional Prices in the future, and to share information that is not publicly available about those prices, both of which are alleged here, that has a chilling effect on competition.³⁴ Further, if Dr. Stiroh's argument was correct, then none of Defendants' pervasive, contemporaneous documents expressing concerns about "bidding wars" would make any sense.³⁵ Similarly, the Overlap Group would have made no economic sense based on her argument, nor would the rationale for the 568 exemption itself.

Furthermore, the quantitative evidence that Dr. Stiroh provides in support of her argument is based on comparing EFCs with faulty benchmarks that would themselves already be inflated by the Challenged Conduct, including the IM.³⁶ I respond to Dr. Stiroh's arguments, among other ways, by demonstrating via regression analysis that the Challenged Conduct resulted in an artificial inflation

33. Stiroh Report ¶70 ("[T]here is no economic nexus between an alleged agreement on certain principles, starting points, and approaches for assessing family ability to pay and the ultimate amount of aid any student was offered, or the net price charged by any Defendant school to any proposed Class member.").

34. See, e.g., Yu Awaya & Vijay Krishna, *Information Exchange in Cartels*, 51(2) RAND JOURNAL OF ECONOMICS 421-446, 421 (2020) ("Antitrust authorities view the exchange of information among firms regarding costs, prices, or sales as anticompetitive. Such exchanges allow competitors to closely monitor each other, thereby facilitating collusion."); *Information Exchanges between Competitors under Competition Law*, OECD DIRECTORATE FOR FINANCIAL AND ENTERPRISE AFFAIRS COMPETITION COMMITTEE: GERMANY (Jul. 11, 2011), https://www.oecd.org/content/dam/oecd/en/publications/reports/2011/07/information-exchanges-between-competitors-under-competition-law_bd644d8b/327f7dd3-en.pdf ("Generally speaking, in a concentrated market the exchange of confidential company data restrains competition if the data is attributable to individual competitors or particular business transactions. Regular meetings combined with up-to-date and individualized data enable the organizers to monitor the participants' behavior closely and to reduce the likelihood of competitive advances. The restriction of access to the circle and the exclusion of smaller competitors can lead to anticompetitive foreclosure and may therefore violate competition rules."); Margaret C. Levenstein & Valerie Y. Suslow, *What Determines Cartel Success?*, 44(1) JOURNAL OF ECONOMIC LITERATURE 43-95, 44 (2006) ("Cartels much prefer to develop the means to monitor each other's behavior in order to deter or physically prevent cheating, rather than resorting to expensive punishments such as price wars. Designing effective monitoring mechanisms takes place over time as cartels learn about both their competitors and their customers, and then refine the organizational structure to provide the necessary incentives and information to sustain cooperation"); *id.* at 71 citing David Genesove & Wallace P. Mullin, *Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890-1914*, 29(2) RAND JOURNAL OF ECONOMICS 355-377 (1998) ("Genesove and Mullin (2001) provide a rich and detailed illustration of the cartel learning process in their analysis of the evolution of agreements among U.S. sugar producers. They describe how weekly meetings of sugar refiners allowed them to 'complete the contract' by adjusting the agreement to changing external conditions, such as fluctuations in demand, and by addressing issues originally left ambiguous."); ADAM SMITH, AN INQUIRY INTO THE WEALTH OF NATIONS, 105-106 (MetaLibri 2007) ("People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices.").

35. See, e.g., Singer Report n. 130 (citing DARTMOUTH_0000359527), n. 291 (citing VANDERBILT-00047310), n. 412 (citing DARTMOUTH_0000359371 at -531). See also *id.* ¶221; Deposition of Patricia McWade (Dec. 8, 2023) [hereafter McWade Dep.] 190:3-191:4 (568 members "wanted to see if we could agree on how to analyze families' ability to pay," and were concerned that "after the exemption expired, oh, we're going back to the Wild Wild West where schools would do . . . what they wanted and different things."). See also, e.g., Appendix 3; Part III.A.1; Singer Report Appendix 7; Singer Report II.D.

36. See Part III.A.1.

in EFCs, contrary to Dr. Stiroh's assertion. I also explain another mechanism by which the Challenged Conduct had artificially inflated Effective Institutional Prices, via the implementation of the affordability principle on packaging—an argument that Dr. Stiroh dismisses entirely, and without basis, as I explain further below.³⁷ I pointed out in my Initial Report, the Challenged Conduct included multiple other prongs, including the sharing of competitively sensitive information, organizational elements that would have effectuated supracompetitive pricing, and inducements to support collusion, such as through regular meetings and surveys.³⁸ These prongs would also have contributed to artificially inflated Effective Institutional Prices.³⁹

Defendants' Experts raise numerous critiques of my overcharge regressions that I present as quantitative evidence of the first step of my two-step methodology to proving common impact.

Dr. Stiroh claims that my use of a common overcharge across Defendants and academic years is economically implausible.⁴⁰ Dr. Stiroh's critiques are incorrect, as explained in more detail in Part III.A.2.a below. My use of a common overcharge is widespread in economics, and it is necessary to reliably measure the overcharge in the instant case because many Defendants' data only cover a limited period in which they had not participated in the Challenged Conduct. Furthermore, over half of Defendants did not produce any pre-Challenged Conduct data, rendering it less likely for one to observe the artificial overcharge due to the umbrella effect. Dr. Stiroh's Defendant-specific regressions are also flawed because they only compare Effective Institutional Prices for a single Defendant during years they had participated relative to years they had not participated in the Challenged Conduct. My use of a common overcharge also accounts for Effective Institutional Price variation between participating and non-participating Defendants during the same academic year. Also, there is no *a priori* basis to allow the conduct variable to vary by time period, contrary to Dr. Stiroh's methodology presented in Part VII.A.2 of her report.

Dr. Stiroh asserts that elements of the Challenged Conduct are present in my regression "benchmark" periods, rendering my regressions incapable of measuring impact.⁴¹ Dr. Stiroh's argument implies that my regressions *underestimate* the artificial overcharge in Effective Institutional Prices resulting from the Challenged Conduct.

Dr. Hill critiques my standard errors, and he claims that his "corrected" standard errors materially weaken my results.⁴² Dr. Hill's replacement of my robust

37. Stiroh Report ¶114.

38. Singer Report §II.D.2.

39. *Id.* §II.D.1.

40. *Id.* §VI.A.

41. *Id.* §VI.B.

42. Hill Report §8.1.

standard errors with two-way clustered standard errors has no bearing on my overcharge estimate; it only artificially decreases the likelihood of obtaining statistical significance by effectively throwing away data. Dr. Hill’s rationale for his standard error replacement is baseless—whether one should cluster or not depends on the level of the treatment assignment, which is not randomly assigned here.

Dr. Hill enumerates numerous purported data-processing and control-variable construction “errors,” and he erringly claims that “correcting” these “errors” results in my regressions showing no reliable evidence of overcharges.⁴³ From the outset, Dr. Hill’s latter assertion is unfounded—he misconstrues the fact that his coefficient estimates are not statistically significant with them not being economically significant. Put differently, he still gets a positive conduct coefficient of \$441 using my primary regression model, but his estimate is not statistically significant when using his two-way clustered standard errors. Correcting his standard errors results in both a positive and statistically significant overcharge using my primary regression model. Furthermore, I agree with some of Dr. Hill’s data-processing and control-variable-construction adjustments, but I disagree with others. I find that his results are highly sensitive to only three adjustments—his inclusion of flawed Chicago post-2015 data, his use of incorrect Duke awards variables, and his aggressive replacement of purported “outlier” values. In Part III.A.2.c, I explain why there is no basis for these three changes, and I find that correcting them results in my primary regression model’s conduct coefficient equaling 1,202—significantly higher than Dr. Hill’s invalid estimate in a statistically significant way.

Dr. Hill asserts that I fail to control for rising costs and macroeconomic conditions that may affect Effective Institutional Prices, and that these factors may thereby result in omitted variable bias.⁴⁴ His arguments are inapposite, most importantly because *Dr. Hill incorrectly defines omitted variable bias*, claiming it could be due to factors “unrelated to the Challenged Conduct.”⁴⁵ Putting aside this misconstrued definition, Dr. Hill’s arguments are still invalid because they ignore that I included *real* prices (deflated using the CPI) and control variables for trend, unemployment, real GDP, and a COVID-19 variable in my regression models, all of which account for rising costs and macroeconomic factors.

Dr. Hill claims that I should have omitted “flawed” post-Class Period data. His argument is unfounded—he does not provide any cite in support of why limited data should be ignored if the researcher cannot obtain complete data. Additionally, his flawed argument also ignores the fact that Defendants were requested to produce complete 2023–2024 data, and that these data are of

43. *Id.* §8.2.

44. *Id.* §8.3. Bias is defined as “[t]he difference between the expected value of an estimator and the population value that the estimator is supposed to be estimating.” Put differently, when I refer to bias in this report, I am referring to whether the point estimate of a regression coefficient equates to its actual (population) value. *See, e.g.,* WOOLDRIDGE at 845.

45. Hill Report §8.3.

particular importance given the lack of variation in the Challenged Conduct period.

Dr. Hill asserts that my student fixed effects models are conceptually flawed.⁴⁶ He asserts that Emory and Dartmouth do not inform the conduct coefficient in these regressions because they only produced one year of data per Class Member. It is my understanding that Defendants were asked to produce *complete* structured financial aid data for this case; hence, this argument reflects a misleading attempt to deflect the responsibility from the Defendants to me. Further, my non-student fixed effects, which do use Emory and Dartmouth data for identification, show an economically and statistically significant overcharge. Dr. Hill also asserts that my student fixed effects models primarily only capture Effective Institutional Price variation for locked-in students. This argument is irrelevant—if anything, this would mean that my regressions underestimate the effect of the Challenged Conduct on Effective Institutional Prices. My use of student fixed effects is a widely used technique throughout the econometrics literature for controlling for time invariant differences between students, and these models are therefore not “conceptually flawed.”

Dr. Hill’s also claims that my models suffer from selection bias. His argument ignores the purpose of my analysis—to measure the overcharge *specific to Class Members*. By limiting the data to Class Members, I do precisely that.

Step 2: In the second step of my two-step approach, I employed two different means of demonstrating that the generalized overcharge in Effective Institutional Prices had resulted in widespread impact across the Class. I first performed a standard in-sample prediction methodology to demonstrate that the general Effective Institutional Price increase established in the first step caused all or nearly all members of the Class to pay artificially inflated prices, and thus suffer impact.⁴⁷ I then used a standard economic price structure analysis to demonstrate that Defendants’ prices tended to move together and are linked such that a change to the Effective Institutional Price to one Class Member would be linked with a change in the Effective Institutional Price to another Class Member.⁴⁸ I also provided qualitative evidence that Defendants’ goals of maintaining horizontal equity is consistent with classwide impact.⁴⁹

Dr. Stiroh is the only Defendants’ Expert to offer an opinion on this second step of my two-step approach to demonstrating common impact. Dr. Stiroh does not acknowledge nor dispute the qualitative evidence that I brought forward regarding Defendants’ goals of maintaining horizontal equity. Instead, she argues that common impact could not have occurred because each separate component of the Challenged Conduct, standing alone, would not have produced

46. *Id.* §8.4.

47. Singer Report §III.B.1.

48. *Id.* §III.B.2.a.

49. *Id.* §III.B.2.b.

classwide impact.⁵⁰ Dr. Stiroh's approach is misguided because one does not need to assess whether each *separate* component of the Challenged Conduct would have harmed all or nearly all Class Members; rather, because Plaintiffs are contesting the Challenged Conduct as a whole, the focus is on whether that the Challenged Conduct *as a whole* resulted in classwide impact.

Dr. Stiroh claims that my in-sample prediction methodology is erroneous because it "finds impact where there is none" by "conflat[ing] the variation in outcomes unexplained by his model with average impact."⁵¹ Her argument is without merit. It is predicated on obtaining a zero overcharge in the first step of my two-step methodology. That my first step shows a generalized overcharge renders her argument invalid. Dr. Stiroh claims that my price structure regressions are unreliable because market-wide factors could result in my regressions showing correlation where there is none.⁵² Her assertion ignores the fact that my price structure regressions control for myriad factors that shift market-wide prices.

Dr. Hill does not address the second step of my two-step methodology to demonstrating common impact. Critically, though, all of his "corrections" involve classwide adjustments. None requires any individualized inquiry. For instance, Dr. Hill suggests using clustered errors, dichotomizing continuous variables, and applying cost indices particular to the higher education industry. Even if he were correct (and he is not), all of these criticisms reveal his implicit agreement that harm can be shown based on a classwide formulaic basis. He only critiques various aspects of my model with which he disagrees. But he implicitly concedes that regression modelling is the appropriate tool to use to show impact and damages here. Indeed, he draws classwide conclusions based on his own alternative analysis using less granular IPEDS data than that which I used.

Put together, the two steps of my common-impact methodology demonstrate based on classwide evidence and analyses that the Challenged Conduct resulted in an artificial generalized overcharge in Effective Institutional Prices, and that this generalized overcharge harmed all or nearly all Class Members. For the foregoing reasons, I find that classwide evidence is capable of proving, and does prove, a common impact to Class Members in the form of higher Effective Institutional Prices paid to Defendants.

- (4) **Aggregate Damages.** In my Initial Report, I used common methods and evidence to quantify aggregate damages to the Class resulting from the Challenged Conduct.⁵³

50. Stiroh Report ¶81.

51. *Id.* ¶197.

52. *Id.* ¶204.

53. Singer Report §IV.

None of Defendants' Experts critiques my methodology for calculating aggregate damages. I find that classwide evidence is capable of establishing aggregate damages owed to the Class. After making the particular data-processing and control variable construction adjustments described in the common impact section, I calculate updated aggregate damages to Class Members of \$685 million.

3. In Appendix 2, I describe and respond to qualitative evidence brought forward by Dr. Hill with respect to Defendants' consensus on financial aid policies. I find that Dr. Hill's presentation of qualitative evidence relies on out-of-context statements. In Appendix 3, I respond to Defendants' Experts various "prepackaged" analyses, which are analyses brought forward that do not respond to my Initial Report, but rather attempt to present a new story. I refer to these analyses as "prepackaged" analyses because they are unresponsive to my own analyses, and they did not require seeing my Initial Report to create. In Appendix 4, I present a rebuttal analysis to Dr. Long's assertion that Defendants are not revenue-maximizing using admissions data from Cornell, Georgetown, MIT, Notre Dame, and Penn.⁵⁴ I find that Dr. Long's opinion is inconsistent with the fact that these Defendants admitted "priority-designated" applicants at higher rates than non-priority designated applicants, even though priority-designated admits include donor-related admits and generally did not score higher on standardized tests. Dr. Long's opinion is also at odds with Dr. Stiroh's assertion that "competitive forces" discipline Effective Institutional Prices.⁵⁵

BACKGROUND

4. In my Initial Report, I described the nuances of higher educational institution pricing relative to other markets for purposes of exposition. In particular, higher educational institutions price discriminate through the provision of need-based and merit-based aid.⁵⁶ I explained the use of EFC

54. Long Report ¶22.

55. Stiroh Report ¶140.

56. Singer Report ¶¶25-30.

formulae by institutions as a means for determining families' ability to pay.⁵⁷ I then described the advent of the 568 Group and its adoption of the CM for determining EFC.⁵⁸

5. I do not offer any affirmative opinions regarding higher educational institutional background. Therefore, I treat any opinions that Defendants' Experts proffer with regard to background as ancillary to the four major opinions summarized in the introduction above. For this reason, any responses I provide pertaining to Defendants' Experts regarding background are presented in Appendix 2.

I. COLLECTIVE MARKET POWER

6. In my Initial Report, I demonstrated that Defendants collectively enjoyed pricing power over their Effective Institutional Prices.⁵⁹ I presented both direct evidence and indirect evidence. Direct evidence consists of my regression results that Defendants were able to artificially inflate Effective Institutional Prices above competitive levels.⁶⁰ It would not have been possible for Defendants to artificially inflate prices unless they collectively held market power over Class Members in the first place. For indirect evidence, I first showed that educational services offered by Elite Private Universities, which I define as private universities ranked in the top 25 of the USNWR rankings from 2003-2022, constitutes a relevant antitrust product market.⁶¹ I applied the HMT to the Elite Private Universities market, and this test showed that Defendants were able to artificially inflate Effective Institutional Prices by an economically significant degree over decades.⁶² I also assessed *Brown Shoe* factors, such as (1) qualitative evidence showing industry and public recognition of the

57. *Id.* ¶¶31-45.

58. *Id.* ¶¶46-58.

59. *Id.* §I.

60. *Id.* §I.A.

61. *Id.* ¶60.

62. *Id.* §I.B.1.b.

USNWR rankings,⁶³ (2) a peer analysis which showed that Defendants perceive undergraduate services provided in this market as separate from other markets,⁶⁴ (3) a travel distance analysis which showed that students view Elite Private Universities distinct from other universities,⁶⁵ and (4) a revealed preference ranking analysis which showed that students generally rank Elite Private Universities higher than other universities.⁶⁶ Next, I showed that Defendants collectively held high shares in the Elite Private Universities Market from 2003-2022.⁶⁷ Lastly, I demonstrated that there are high barriers to entry in the Elite Private Universities Market.⁶⁸

A. Dr. Hill’s Critique of My Hypothetical Monopolist Test Is Invalidated by My Regression Results

7. My Initial Report provides direct evidence that Defendants were able to artificially inflate Effective Institutional Prices during the Class Period by an economically and statistically significant amount.⁶⁹ Dr. Hill states that he finds no evidence of Defendants artificially inflating Effective Institutional Prices after he adjusts my “data processing and modeling errors[,]” and that his findings therefore contradict my direct evidence.⁷⁰

8. I disagree with numerous adjustments that Dr. Hill makes to my data processing and modeling, and I find that once I correct for his improper adjustments, my regressions continue to show a positive, economically, and statistically significant artificial overcharge in Class Members’ effective institutional prices due to the Challenged Conduct. I explain these findings in more detail in Part III.A.2 below. This evidence of an artificial overcharge during the Class Period demonstrate

63. *Id.* §I.B.1.c.

64. *Id.* §I.B.1.d.

65. *Id.* §I.B.1.e.

66. *Id.* §I.B.1 f.

67. *Id.* §I.B.2.

68. *Id.* §I.B.3.

69. Singer Report ¶63.

70. Hill Report ¶159.

that Defendants were collectively able to profitably raise effective institutional prices to Class Members above competitive levels over many years, contrary to Dr. Hill's opinion.

B. Defendants' Experts Claim That Defining the Relevant Market As a Subset of Universities Ranked Highly by USNWR Is Not Appropriate

1. Dr. Hill and Dr. Long Incorrectly Claim USNWR Rankings Are the Basis for My Market Definition

9. In my Initial Report, I defined the relevant market as private universities consistently ranked highly by USNWR. Dr. Hill and Dr. Long assert that I used USNWR as the "basis" to form my market definition.⁷¹ This is inaccurate—I provided evidence for multiple different *Brown Shoe* factors to inform my market definition, and these factors provide practical indicia supporting the relevant market as defined. This evidence includes (i) qualitative evidence of industry and public recognition, (ii) a peers analysis showing these institutions regard themselves as distinct from institutions outside the relevant market, (iii) an analysis showing the relevant market universities have a largely nationally representative student body, more so than other institutions, and (iv) pairwise comparisons of enrolment decisions showing undergraduates having a clear preference for enrolling in relevant market schools when also having a non-relevant market option available.⁷² The USNWR rankings are an important consideration, but they are not "the basis" of my opinion.

2. Dr. Hill Incorrectly Throws out Any Consideration of the USNWR Rankings for Market Definition

10. Dr. Hill argues that any consideration of the USNWR Rankings for a market definition is flawed.⁷³ He is wrong. The partial reliance on USNWR rankings for defining the market is justified by the empirical analysis of the practical indicia outlined above. Further, qualitative evidence shows the import of USNWR rankings for the first practical indicia of "industry or public recognition,"

71. *Id.* ¶27. Long Report ¶26.

72. Singer Report §I.B.1.c-f.

73. Hill Report ¶150.

which is to say that both schools and students pay them considerable attention. *The New York Times* recently reported that “to students and their parents, the rankings can be tools for narrowing college searches, and status symbols surrounding admissions to certain schools. To university leaders, the rankings are often publicly heralded.”⁷⁴ Indeed, when USNWR released the latest ranking in September 2024, the majority of Defendants published press releases discussing and often touting their respective rankings.⁷⁵

11. Further, Defendants’ documents show they used USNWR rankings to assess their own competitive position for the purposes of admissions. Notre Dame’s President Rev. Edward A. Malloy has stated that Notre Dame considered USNWR top 20 schools as its competition.⁷⁶ When asked about a section of a 2019 draft document titled “Peer enrollment strategy summary (USNWR Top 20),” Malloy testified about the USNWR Top 20 schools listed in the document, “they’re among our

74. Alan Blinder, *The U.S. News College Rankings Are Out. Cue the Rage and Obsession*, NEW YORK TIMES (Sept. 24, 2024), <https://www.nytimes.com/2024/09/24/us-us-news-rankings-colleges.html>.

75. Johns Hopkins Rises To No. 6 In ‘U.S. News’ Best Colleges Rankings, JOHNS HOPKINS UNIVERSITY HUB (Sept. 24, 2024), <https://hub.jhu.edu/2024/09/24/us-news-best-colleges-rankings-2024/>. Cate Latimer, *Brown University drops from Top 10 in 2025 U.S. News Ranking*, THE BROWN DAILY HERALD (Sept. 24, 2024), <https://www.browndailyherald.com/article/2024/09/brown-university-drops-from-top-10-in-2025-u-s-news-ranking>. Abby Spiller, *Duke rises to No. 6 in U.S. News and World Report national ranking, highest in 19 years*, THE CHRONICLE (Sept. 23, 2024), <https://www.dukechronicle.com/article/2024/09/duke-university-sixth-us-news-and-world-report-national-university-ranking-america-highest-since-2006-top-10-caltech-johns-hopkins-northwestern>. Nineth Kanieski Koso, *Northwestern reaches historic high at 6th place in U.S. News college rankings*, THE DAILY NORTHWESTERN (Sept. 24, 2024), <https://dailynorthwestern.com/2024/09/24/campus/northwestern-reaches-historic-high-at-6th-place-in-u-s-news-college-rankings/>. Anushka Shorewala, *U.S. News Ranks Cornell No. 11 University in Country, Best in New York*, THE CORNELL DAILY SUN (Sept. 23, 2024), <https://cornellsun.com/2024/09/23/cornell-ranked-no-11-university-in-country-best-in-new-york-in-u-s-news-and-world-report/>. MIT named No. 2 university by U.S. News for 2024-25, MIT NEWS (Sept. 24, 2024), <https://news.mit.edu/2024/mit-named-no-2-university-us-news-0924>. Spencer Davis, *Columbia ranks No. 13 in U.S. News rankings, falling one spot*, COLUMBIA SPECTATOR (Sept. 24, 2024), <https://www.columbiaspectator.com/news/2024/09/25/columbia-ranks-no-13-in-us-news-rankings-falling-one-spot/>. Laura Diamond, *U.S. News names Emory among top national universities*, EMORY NEWS CENTER (Sept. 24, 2024), https://news.emory.edu/stories/2024/09/er_us_news_undergraduate_rankings_24-09-2024/story.html. Mia Streitberger, *Georgetown Remains at No. 22 in U.S. News 2024 Top Colleges List*, THE HOYA (Sept. 23, 2024), <https://thehoya.com/news/georgetown-remains-at-no-22-in-u-s-news-2024-top-colleges-list/>. Katie Bartlett, *Penn falls to lowest U.S. News ranking since 1997 while Princeton, MIT notch top spots*, THE DAILY PENNSYLVANIAN (Sept. 10, 2024), <https://www.thedp.com/article/2024/09/penn-princeton-mit-us-news-rankings-drop>. Chris Stipes, *Rice in top 20 of US News ‘Best Colleges’ rankings*, RICE UNIVERSITY NEWS AND MEDIA RELATIONS (Sept. 24, 2024), <https://news.rice.edu/news/2024/rice-top-20-us-news-best-colleges-rankings-0>.

76. Deposition of Edward Malloy (Sept. 7, 2023) [hereafter Malloy Dep.] 104:13-23. See also ND_0019093 at -118 (slide showing the “Peer Enrollment Strategy Summary” for the “USNWR Top 20.”).

principal competitors.”⁷⁷ When asked about the relevance of the USNWR rankings, he responded that “they increased the percentage of those who applied and the quality that applied.”⁷⁸

12. Emory University assigns so much importance to its USNWR position that it intentionally supplied incorrectly inflated data to USNWR from 2000 to 2011.⁷⁹

13. An April 2014 Cornell document titled “Enrollment Assessment Task Force” contains the following observation: “Although other rankings exist, the U.S. News & World Report rankings arguably garner the most media attention and undoubtedly serve as a guide for students, parents, and guidance counselors in the college search process.”⁸⁰ And the Cornell Department of Economics Chair Amanda Griffith wrote in a peer-reviewed academic publication that: “We find that school choice is responsive to changes in [USNWR] rankThe results indicate that the USNWR rank is an important factor in the decisions of high-ability students.”⁸¹

14. A document titled “Dartmouth’s Competitive Positions” notes: “In the US News ranking—the most widely referenced ranking and the main focus of this paper—Dartmouth has held in ninth place for the past 5 years.”⁸² The document goes on to detail Dartmouth’s competitive position based on its USNWR ranking and various ranking criteria.⁸³ The document also lists Dartmouth’s “close competitors,” which are all in the USNWR top 25.

15. When asked whether moving up in the USNWR rankings was important to Northwestern, the school’s Dean of Undergraduate Enrollment Chris Watson stated: “Yes. I would say yes.”⁸⁴

77. Malloy Dep. 104:13-18.

78. *Id.* 104:19-23.

79. Kevin Kiley, *Sorry, Wrong Numbers, INSIDE HIGHER ED* (Aug. 12, 2012), <https://www.insidehighered.com/news/2012/08/20/emory-misreported-admissions-data-more-decade>.

80. CORNELL_LIT0000004362 at -393.

81. CORNELL_LIT0000214653 at -653.

82. DARTMOUTH_0000158513. *See also* Singer Report ¶101.

83. DARTMOUTH_0000158513.

84. Deposition of Chris Watson (Oct. 10, 2023) 134:4-16.

16. Vanderbilt's Director of Undergraduate Admissions John Gaines testified "U.S. News was important . . . And as director of admissions when I was speaking to prospective families, especially those who were early in the process and fixated on U.S. News, I would remind them . . . look at the fact that Vanderbilt was always there, always there in those rankings somewhere . . . And we used that fact in admissions as evidence that there are all kinds of third parties out there saying that great things are happening at Vanderbilt, and it worked. The students really loved reading those ratings, and we loved promoting them."⁸⁵

17. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]⁶

C. Academic Literature Does Not Undermine My Reliance on USNWR Rankings

1. The Hu Paper Is Not Inconsistent with My Use of the USNWR Rankings.

18. In my Initial Report, I cited a 2020 paper by Xingwei Hu, which shows how to apply a revealed-preference technique to data on students' enrollment decisions to generate a school ranking distinct from that published by USNWR. Dr. Hill states: "Hu (2020) argues that using college rankings to evaluate schools is inferior to analyzing observed student choices."⁸⁷

19. But I do analyze observed student choices in my revealed-preference method, and those results are consistent with USNWR rankings for the top 25 ranked universities. Table 7 of my Initial Report illustrates that this revealed-preference ranking has schools in the relevant market correlated at the top.⁸⁸ Applying Hu's revealed-preference method, which is clearly distinct from the

85. Deposition of John Gaines (Aug. 3, 2023) 207:10-208:7.

86. DUKE568_0124311 at -312. *See also* Michael Thaddeus, *College rankings whistleblower: Exposing inaccurate data was unpleasant but necessary*, CNN (Sept. 22, 2022), <https://www.cnn.com/2022/09/22/opinions/columbia-ranking-inaccurate-data-thaddeus/index.html>.

87. Hill Report ¶153.

88. Singer Report ¶105.

USNWR ranking method, nonetheless produces a similar ranking at the top. The reason I apply Hu's method of analyzing students' revealed preferences is to illustrate that my defined market is robust to the types of flaws in USNWR rankings that concern Hu.

2. The Rankings from the Avery et al. Paper Have Not Displaced the USNWR Rankings

20. The Avery et al. (2013) paper is well regarded in academic circles. Yet the resulting rankings in the paper have certainly not risen to the level of "industry and public recognition" specified in the *Brown Shoe* factors, even though that paper was published in 2013 using data collected in 2004. I apply the Hu revealed-preference technique to more recent parchment.com data to support my defined market.

21. As Dr. Hill states, Avery et al. point out that that the weighted criteria used in rankings such as USNWR can be manipulated by schools.⁸⁹ And, indeed, as outlined in I.B above, several Defendants appear to have possibly engaged in such manipulation. This is further proof that USNWR rankings clearly constitute "industry and public recognition," of the nation's elite schools; there would be no incentive to manipulate one's ranking. And the peers' analysis summarized in Table 3 and Table 4 of my Initial Report show this recognition is concentrated within schools in the relevant market, and less concentrated across relevant market and non-relevant market schools.

3. Dr. Hill Claims That Applying Hu's Revealed Preference Method Using Parchment.com Data Are Inconsistent with the Market Definition

22. Dr. Hill claims that once I add ten additional universities to my revealed-preference analysis, it reveals "significant competition between private schools and public schools."⁹⁰ But Table 7 of my Initial Report does not measure the strength of competition between pairings of schools. Rather, it shows that, among this grouping of 32 schools, the 22 schools in the relevant market

89. Hill Report ¶154.

90. *Id.* ¶160.

account for 87.7 percent of the total impact or “authority” per the revealed-preference analysis, and the additional ten schools outside the relevant market account for the remaining 12.3 percent. Limited competition for students from these ten outside schools is not in and of itself an issue, just as application of the hypothetical monopolist test (HMT) reveals *some* customers defect from the hypothetical monopoly is not in and of itself an issue. It is only an issue if the level of defection is sufficiently high for a price rise to cause a fall in net assets (i.e., “profits”).⁹¹

23. I therefore find that the revealed-preference evidence from parchment.com data complements the direct evidence in my Initial Report, suggesting that Defendants collectively held market power to be able to raise Effective Institutional Prices above competitive levels.

4. Dr. Long’s Claim That Some Students Applying to a Wide Range of Schools Implies a Broader Relevant Market Here Is Wrong

24. Dr. Long highlights that some students may simultaneously apply to public universities, liberal arts colleges, and lower-ranked private national universities alongside the schools in the relevant market.⁹² This argument is without merit because the application decision is very different from the *enrollment* decision. Students apply to a wide range of schools not knowing where they will get in. Indeed, some schools serve as an insurance policy for gifted applicants; they would never go to the fall-back school conditional on being accepted at one of their targets. The relevant market here, in contrast, is where students decided to *enroll*, once they know their options.

25. To illustrate my reasoning, consider a student who applies and is accepted to both Harvard University and Rutgers University. Generally, a student receiving offers from both

91. Defendants are all organized as non-profit enterprises. Accordingly, they do not distribute profits to shareholders. Instead, “profits” in this instance can be thought of as a change in net assets, since the residual funds remaining after paying out all expenses are generally reinvested back into the school or into its endowment. *See* Jo-Anne Williams Barnes, *Do Nonprofit Organizations Have Profit and Loss Statements?*, JFW ACCOUNTING SERVICES (Oct. 10, 2022) (“The statement of activities is a nonprofit’s organization income statement. While a traditional income statement exists to show a profit, or Net Income, a statement of activities exists to show the change in net assets.”).

92. Long Report ¶26.

universities is going to be more inclined to attend Harvard over Rutgers based on both universities' yield rates.⁹³ That both Harvard and Rutgers accepted the student is irrelevant; what ultimately matters for purposes of defining the relevant market is where the student chooses to enroll. A similar logic applies when one conducts an HMT. For instance, a given customer may choose to no longer purchase from some firm if the firm chooses to increase prices by an economically significant amount over a non-transitory period of time. That the firm may lose some customers due to the price increase is not relevant; what matters for purposes of the HMT is the *aggregate* loss of customers.

D. Defendants' Experts Are Wrong That Public Schools Belong In The Relevant Market

1. Dr. Hill Incorrectly Cites the Application Behavior of Certain Named Plaintiffs as Evidence That Public Universities Belong in the Relevant Market

26. Dr. Hill cites the fact that six of the eight named Plaintiffs applied to both a public and a private school as evidence of significant competition between private and public schools.⁹⁴ As explained earlier, application decisions are irrelevant. The six plaintiffs in question all chose to *enroll* at Defendant schools, despite being accepted at the public schools. Table 6 of my Initial Report presents further evidence of the distinction between the relevant market and public schools, which shows public schools have considerably less national reach than the private schools comprising the relevant market.

93. For instance, Harvard's "yield rate," which measures the number of accepted students that end up choosing to enroll, was 84 percent for their Class of 2027. See Michelle N. Amponsah and Emma H. Haidar, *84% of Admits Accept Spots in Harvard College Class of 2027*, THE HARVARD CRIMSON (May 20, 2023), <https://www.thecrimson.com/article/2023/5/20/class-of-2027-yield-data/>. Rutgers' yield rate for the Class of 2027 was 27 percent. See Dave Bergman, *How to Get Into Rutgers University: Admissions Data & Strategies*, COLLEGE TRANSITIONS (Jul. 30, 2024), <https://www.collegetransitions.com/blog/how-to-get-into-rutgers-university/>.

94. Hill Report ¶157.

2. Dr. Hill Incorrectly Asserts That Defendants’ Own Documents Show Public Universities Belong in the Relevant Market

27. Dr. Hill also claims that the Defendants’ own documents point to competition with highly ranked public schools.⁹⁵ She is wrong. Many of the Defendants’ documents explicitly downplay public schools as competitors.

28. Dr. Hill relies on a Penn document that includes four tiers of Penn competitors, the last of which is titled “Public Flagships.”⁹⁶ The document ascribes the following descriptive names to the first three categories in the hierarchy: (1) “Top Competitors” (Harvard, Yale, Princeton, MIT, Stanford), (2) “Head-to-Head” (Brown, Columbia, Cornell, Dartmouth, Duke, UChicago), and (3) “Penn Aspiring” (Caltech, Carnegie Mellon, Georgetown, Johns Hopkins, NYU, Northwestern, Rice, USC, Vanderbilt, and WashU). That Penn segregates state schools in a distinct category below the other three tiers demonstrates that Penn did not consider “Public Flagships” as consequential as the others. Further, Dr. Hill ignores other Penn documents, such as when Senior Executive Vice President Craig Carnaroli states in a February 16, 2015 email to Dean of Admissions Eric Furda: “While our dual degree programs help us in competing for the brightest undergraduate applicants against peers like H-Y-P, Stanford and MIT, our principal competition for students includes Columbia, Cornell, Duke, and Chicago.”⁹⁷ Dean Furda responds: “The only schools we ‘lose’ to are H, Y, P, Stanford. Columbia and MIT to a lesser degree, and we are making gains against Columbia. The set you include below are certainly our nearest peers for application competition and US News rankings.”⁹⁸ Public universities are not included among the competitors mentioned here.

95. *Id.* ¶156.

96. PENN568-LIT-00089925 at -951.

97. PENN568-LIT-00004598.

98. *Id.*

29. Dr. Hill fails to consider documents, such as [REDACTED]

[REDACTED]

[REDACTED]⁹⁹

30. Dartmouth listed its “close competitors” as Columbia, Northwestern, Washington University and Brown; there is no public school listed.¹⁰⁰

31. In a 2016 email, Georgetown University’s Dean of Admissions Charles Deacon listed Georgetown’s peers in four tiers. The uppermost tiers list top 20 USNWR schools and include (1) Harvard, Yale, Princeton, Stanford, (2) Brown, Penn, Duke, Columbia, Dartmouth, (3) Cornell, Chicago, Northwestern, Hopkins, Wash U, Vanderbilt, Emory, Notre Dame.¹⁰¹ Nowhere on Georgetown’s list of competitors, not even within its lowest tier, does it include public universities.

32. In conclusion, none of Defendants’ Experts’ critiques undermine my finding that Defendants held collective market power in the Elite Private University services market. I continue to show direct evidence of market power via my Effective Institutional Price regressions which, as demonstrated in Part III.A.2.c below, continue to show a positive, economically and statistically significant artificial overcharge in Effective Institutional Prices charged by Defendants. I find Defendants’ Experts’ critiques of my HMT, my use of USNWR as a starting point for my market definition, and my calculation of Defendants’ collective market shares without merit. Defendants’ Experts do not address the evidence I provide establishing the Elite Private University Market using qualitative evidence under the *Brown Shoe* test and documenting the significant barriers to entry in that market.

99. DUKE568_0098047.

100. DARTMOUTH_0000158513 at -519.

101. GTWNU_0000039045.

II. THE CHALLENGED CONDUCT

33. In my Initial Report, I described the general features of the Challenged Conduct.¹⁰² I explained that the Challenged Conduct is an alleged Overarching Agreement comprised of consensus on six elements, including: (1) six core principles of awarding institutional aid, (2) the primacy of awarding need based aid, (3) the use of the Institutional Methodology (“IM”) as the basis for developing and using the Consensus Methodology (“CM”), (4) the development and use of the CM itself, (5) a manual for exercising Professional Judgment that was to applied to exceptional circumstances, and (6) the sharing of competitively sensitive information (“CSI”).¹⁰³ I then analyzed, through an economic lens, whether the Challenged Conduct is consistent with the allegations of conspiracy and inconsistent with unilateral conduct and unfettered competition in setting Effective Institutional Prices.¹⁰⁴ I presented quantitative evidence—namely, my overcharge regression results—which are consistent with *all* Defendants having engaged in the Challenged Conduct to artificially suppress institutional grant aid (and thereby artificially inflate Effective Institutional Prices).¹⁰⁵ I also presented quantitative evidence consistent with *each* Defendant having participated in the Challenged Conduct to artificially inflate Effective Institutional Prices.¹⁰⁶ This evidence consisted of applying in-sample prediction to each Defendant after having run my primary regression overcharge model; this analysis showed that each Defendant had artificially inflated its Effective Institutional Prices during periods in which it had participated in the Challenged Conduct above where such prices would have been in the absence of the Challenged Conduct. This statistical finding

102. Singer Report §II.A.

103. *Id.* ¶148.

104. *Id.* §II.D.

105. *Id.* §II.B.

106. *Id.* §II.C.

demonstrates that each Defendant's prices are consistent with it having participated in the Challenged Conduct, and inconsistent with it having competed in an unfettered way.

34. Defendants' Experts dispute element (4) of the Challenged Conduct, the consensus on use of the CM for determining EFCs, but they do not dispute the execution of the five other elements. Defendants' Experts did not contest my economic assessment of the qualitative evidence pertaining to any of the other elements of the Challenged Conduct, including the exchange of competitively sensitive information on prices, the organizational elements that effectuated supracompetitive pricing, and inducements to support collusion, such as through regular meetings and surveys.¹⁰⁷ Defendants' Experts assert that the five uncontested elements of the Challenged Conduct were not anticompetitive because they were merely the result of industry standards, and, because of this, Defendants would have engaged in the same actions absent an Overarching Agreement.

35. I explain the flaws in Defendants' Experts' arguments below. Namely, I provide both qualitative and quantitative evidence that is consistent with Defendants having used a common formula for EFCs, in direct contrast to Defendants' Experts' arguments, and also with the Challenged Conduct having otherwise inflated Effective Institutional Prices. I explain that the economics of cartels is inconsistent with Defendants' Experts' theory that Defendants would have engaged in the same conduct absent an Overarching Agreement.

36. It bears noting that I presented Plaintiffs' theory of the case for purposes of exposition and for purposes of assessing whether, *conditional on the existence of the Challenged Conduct*, said conduct is consistent with the alleged overarching conspiracy and inconsistent with competition when viewed through an economic lens, as well as whether the economic facts are consistent with each Defendant's participation in the Challenged Conduct. Put differently, my assignment is predicated

107. Singer Report §II.D.2.

on the assumption that Defendants formed the 568 Group, agreed upon principles to award institutional grant aid, and developed the CM as a group. I use qualitative and quantitative evidence to illustrate that Defendants implemented those principles and the CM in a manner that, in conjunction with other aspects of the Challenged Conduct such as the sharing of CSI, resulted in increased Effective Institutional Prices. I offer no opinion as to the existence of an agreement or to whether any Defendant has violated the antitrust laws.

A. Contrary to Defendants’ Experts’ Arguments, There Is Substantial Evidence That All Defendants Imposed the Consensus Methodology

37. In my Initial Report, I describe the six elements that comprise the Challenged Conduct, as described in the Summary of Conclusions above.¹⁰⁸ As stated above, Defendants’ Experts dispute only one element of the six—the consensus use of the CM (item 4). Defendants’ Experts argue that there was no consensus to use the CM, because Defendants purportedly implemented the CM inconsistently.¹⁰⁹ As explained below, this argument is incorrect because it ignores both the qualitative evidence provided in my Initial Report and quantitative evidence presented in this rebuttal, both of which are consistent with the Challenged Conduct resulting in a convergence of CM EFCs across Defendants during the Class Period. It also is misleading because it presumes that a divergence in EFCs *of any magnitude* refutes sufficient “consensus” for the Challenged Conduct to have anticompetitive effects and a common impact on the Class.

38. As a matter of economics, the Challenged Conduct could have anticompetitive effects even without the CM, given that the other aspects of the Challenged Conduct, could, in and of themselves, lead to anticompetitive effects and a common impact. Moreover, even if 568 Group members deviate from certain elements of the CM, that part of the Challenged Conduct would lead

108. *Id.* §II.A.2.

109. Stiroh Report ¶86, ¶150; Hill Report ¶¶113-114; Long Report ¶224, ¶239.

to artificially inflated prices and a common impact so long as they set EFCs *greater than or equal to* the CM EFC.¹¹⁰ As Georgetown Dean (and Head of the 568 Technical Committee) Patricia McWade clarified, “it was made clear to those wanting to participate in the 568 [Group] that they could NOT use an EFC that was lower than what is calculated using the CA[.]”¹¹¹ If anything, this one-sided ratchet against Class Members’ interests would exacerbate the effect of the Challenged Conduct, not prove that a conspiracy did not exist.

39. Record evidence shows that Defendants understood membership in the 568 Group to bind them closely to the CM. Penn’s Director of Financial Aid explained to the Penn President that some schools left the 568 Group because “they no longer wish[ed] to be limited by the restrictions of the needs analysis consensus document which requires us to apply certain needs analysis assessment constantly amongst all of our schools. (an example of this would be the percentage of the home equity that is used toward the contribution)[.]”¹¹² The same document informed the Penn President that “both Harvard and Princeton never joined because they did not want to limit the usage of their funds.”¹¹³ Columbia highlighted the same point in a PowerPoint presentation discussing “peer comparisons”, where it stated that Princeton was “[n]ever a member of 568 President’s Group -not bound by Consensus Approach”; Yale was “[n]o longer a member of 568 President’s Group -not bound by Consensus Approach”; and Penn was a “[m]ember of 568 President’s Group -bound by Consensus Approach.”¹¹⁴ In an internal May 2011 presentation about making potential changes to its

110. Emory_568Lit_0000577 (“[T]he Committee recommends that the revised bands be included in the Consensus Approach with the understanding that member institutions are free to continue to use the old bands as appropriate for their institutional needs. Using the old bands will generate higher PCs which is consistent with our policy of making across-the-board revisions only if PCs are increased.”). *See also* Singer Report ¶147.

111. GTWNU_0000193182 (McWade Deposition Exhibit 11). Ms. McWade also added that the requirement to have the EFC match the CM or be greater than the CM calculated value “was why HYP felt the need to leave the group; they wanted to offer more “need-based” scholarship aid.” *Id.*

112. PENN568-LIT-00133526.

113. *Id.*

114. Columbia_00005800.

IM to free up budget, Columbia assumed that the parent contribution formula Columbia used could not be changed unless 568 Group changed the CM.¹¹⁵ Columbia further stated it would have to consider the “implications” for its membership in the 568 Group if it wanted to otherwise modify the formula it used, and “weigh the tradeoffs” of leaving the 568 Group.¹¹⁶ Overall Columbia knew that “[m]embership in the 568 Presidents’ Group commits us to a certain number of needs analysis principles which determine parental contribution (PC).”¹¹⁷

40. The 568 Group additionally limited the flexibility that members had in packaging their financial aid awards to such a degree that Emory’s Office of Undergraduate Education Faculty Advisory Committee “recommend[ed] withdrawing from the 568 Group but retaining need blind admissions” because it would “give Emory far greater flexibility in the packaging of financial aid.”¹¹⁸ The 568 Group thereby not only coordinated the amount of financial aid offered, but also the type of aid offered.

41. Furthermore, inconsistent adoption on some elements of the Challenged Conduct, if true, does not mean that the Challenged Conduct is not consistent with conspiracy. Dr. Long’s argument that different Defendants chose to follow different components of the CM ignores the fundamental threshold fact that financial aid decisions, specifically EFC, are the product of the available income and assets multiplied by their respective assessment rates. These assessment rates, in turn, came from the Base IM that the Defendants agreed to apply as the starting point for the setting of institutional grant aid awards.¹¹⁹ Once a university begins with the Base IM as the starting point,

115. Columbia_00179751 at -755.

116. *Id.* at -756.

117. Columbia_00016697 (also stating that Columbia has “already taken steps to be less generous than many of our peers.”).

118. Emory_568Lit_0006213.

119. Singer Report ¶39, ¶¶42-43, ¶132, ¶148.

it is difficult to change the assessment rates.¹²⁰ Defendants thus effectively agreed to calculate ability to pay by multiplying the same range of assessment rates (that is, the same fractions), based on the same (or very similar income bands), by the calculated available income and assets. This is an agreement on a formula. Neither Dr. Long nor any of the other Defendant Experts address assessment rates or income bands, which underpin the calculations of institutional aid.

42. Moreover, Dr. Long's approach in arguing that Defendants do not apply the CM uniformly is inherently misleading. Dr. Long determined a school as compliant with eight "actionable" CM elements if it was "consistent with all components" for "all years" for which data are available.¹²¹ This approach is inconsistent with the CM Agreement itself. The qualitative evidence indicates that by Defendants' own agreement, compliance with the CM did not require rigid adherence to the guidelines, but merely set a *floor* for determining EFC.¹²² This approach also does not even purport to address the materiality of the deviation substantively or temporally. For instance, to consider just one CM element for one school, Dr. Long determined that MIT purportedly did not follow the guidelines regarding home equity because MIT had an exemption from considering home equity for families with an annual income under \$100,000.¹²³ But the record evidence shows that MIT's policy was that MIT actually would, in appropriate circumstances, consider home equity when it was significant, even for families earning less than \$100,000.¹²⁴ It appears difficult to determine then, strictly from the qualitative evidence, to what extent MIT materially deviated from the CM

120. Deposition of Karen Cooper (Mar. 27, 2024) 105:13-21 ("Q. And with respect to the income assessment rates, the actual rates used, do you know if there are any other schools that use different rates than those set forth in the tables? Mr. Bredar: Objection. A. Yeah, not that I'm aware of. It would be difficult to do. You'd basically have to write your own software.").

121. Long Report ¶177.

122. See, e.g., DARTMOUTH_0000359371 at -379; GTWNU_0000193182; MIDDLEBURY001293. The ability to deviate "up" from the guidelines is also why the qualitative evidence compiled in Stiroh Report Ex. 4 regarding the Defendants' backward-looking testimony on whether Defendants felt bound to the CM is not inconsistency with the Challenged Conduct.

123. Long Report ¶177.

124. MITLIT-000169873 at -933.

guidelines regarding home equity. My EFC regression analysis below in Table 1 addresses this inconclusive qualitative information by showing that the Challenged Conduct is indeed associated with an economically and statistically significant increase in the EFC.

43. Defendants' Experts take their argument one step further by claiming that if Defendants were coordinating on elements of the Challenged Conduct, then one would expect to see EFCs and net prices match *exactly* for the same students at different institutions.¹²⁵ Defendants' Experts conduct numerous "prepackaged" analyses (for which they did not need to view my Initial Report to create) to demonstrate that EFCs and net prices do not always match precisely. I describe the flaws of these analyses in detail in Appendix 3. Dr. Hill provides qualitative evidence addressing the argument that Defendants differ in how they calculate financial aid. Much of this qualitative evidence is taken out of context, and when examined in full and through an economic lens, supports the notion of coordination in determining financial aid. I address this out-of-context evidence in Appendix 2.

44. To evaluate how EFCs changed in coordination with the Challenged Conduct, I directly test whether Defendants' EFCs were affected by Defendants' participation in the 568 Group using regression analysis. To do this, I apply the same model that I used to measure the artificial inflation in Effective Institutional Prices owing to the Challenged Conduct (shown in Table 11 of my Initial Report), but I replace the dependent variable Effective Institutional Price with EFC. In Table 1, I report regression results from this EFC conduct regression model. I show all six specifications, consistent with the six specifications of my Effective Institutional Price regression that I had

125. See, e.g., Stiroh Report ¶159.

presented in Table 11 of my Initial Report.¹²⁶ Columns 1-3 show regression results when using Defendant fixed effects, whereas Columns 4-6 show results using student-Defendant fixed effects. The conduct coefficient in each regression shows the artificial inflation in Defendants' EFCs that resulted from their participation in the Challenged Conduct. My primary regression model in Column 6 shows a conduct coefficient of 515, implying that the Challenged Conduct resulted in an artificial overcharge to EFCs of roughly \$515 per student and academic year. This overcharge is statistically significant at a one percent significance level.

126. Because EFC formulae generally assess student factors in levels, I only run EFC regressions in levels. For instance, *see, e.g., The EFC Formula, 2014-2015*, U.S. Department of Education, Federal Student Aid, <https://fsapartners.ed.gov/sites/default/files/attachments/efcformulaguide/091913EFCFormulaGuide1415.pdf> (last visited Oct. 2024) (showing an additional \$4,820 income protection allowance for a family of four versus a family of five, with one in college); *id.* at 10 (showing a student's income assessed at 0.5 and assets assessed at 0.2).

TABLE 1: EFC REGRESSION RESULTS

	Dependent Variable: <i>Real EFC</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
	<i>No Student Fixed Effects</i>			<i>Includes Student Fixed Effects</i>		
Conduct	699.78*** (0.000)	453.71*** (0.000)	753.08*** (0.000)	459.21*** (0.000)	560.73*** (0.000)	514.94*** (0.000)
Adj. Gross Income	89.38*** (0.000)	89.08*** (0.000)	88.88*** (0.000)	51.46*** (0.000)	51.42*** (0.000)	51.50*** (0.000)
Net Worth	2.87*** (0.000)	2.84*** (0.000)	2.82*** (0.000)	0.77*** (0.002)	0.77*** (0.002)	0.77*** (0.002)
Number in College	-4,279.04*** (0.000)	-4,224.73*** (0.000)	-4,221.93*** (0.000)	-6,627.13*** (0.000)	-6,630.58*** (0.000)	-6,629.32*** (0.000)
Year in College	-1,375.72*** (0.000)	-1,423.53*** (0.000)	-1,450.95*** (0.000)	-66.42*** (0.000)	-67.38*** (0.000)	815.94*** (0.000)
Student's Gift Aid from Other Sources	-0.61*** (0.000)	-0.60*** (0.000)	-0.59*** (0.000)	-0.09*** (0.000)	-0.09*** (0.000)	-0.09*** (0.000)
Lagged Excess Endowment Investment Returns		2,057.41*** (0.000)	-1,550.80*** (0.000)		1,264.32*** (0.000)	733.52*** (0.007)
Inst. Tuition Rev per FTE Undergraduate (1-year lag)		0.09*** (0.000)	-0.01*** (0.000)		-0.01*** (0.000)	-0.02*** (0.000)
% of FY-FTE Undergrads Receiving Financial Aid		-21.50*** (0.000)	1.96 (0.757)		-21.68*** (0.000)	-8.77** (0.018)
Unemployment (1-year lag)			-79.66*** (0.000)			-89.51*** (0.000)
COVID			741.01*** (0.000)			-85.00 (0.429)
Trend			34.78 (0.300)			-1,229.23*** (0.000)
Real GDP			0.87*** (0.000)			1.00*** (0.000)
Observations	690,236	690,236	690,236	690,236	690,236	690,236
R-Squared	0.23	0.23	0.23	0.90	0.90	0.90
Includes Institution Fixed Effects?	Y	Y	Y	Y	Y	Y
Includes Institution*Student Fixed Effects?	N	N	N	Y	Y	Y
Number of Fixed Effects	17	17	17	256,661	256,661	256,661

Notes: Robust p-values in parentheses; ***p<0.01, **p<0.05, *p<0.1. Adjusted Gross Income and Net Worth reported in thousands. All dollar denominated variables are in real dollars. These regressions are equivalent to the regressions that I provided in Table 11 of my Initial Report, but using revised data adjusted to incorporate all of Dr. Hill's data-processing adjustments marked as "Accepted" = "Y" outlined in Table 5.

45. Next, to evaluate how the Challenged Conduct impacts the difference in the EFCs offered to cross-admitted students at various Defendants, I analyze the EFCs for cross-admitted students, comparing the two EFCs offered by each Defendant. I evaluate how the range in EFC values is affected by the Challenged Conduct. Compared to Defendants' Experts' analyses, my analysis evaluates how the Challenged Conduct affects the range of EFC values, rather than naively requiring that EFCs match perfectly during the Challenged Conduct to show an effect. I control for the student's average income and net worth (the average of the values as assessed by the two schools), the number in college (also averaged between the two schools), the number of Defendants that admitted the given

student, the difference in U.S. News & World Report rankings for the Defendants that admitted the student, the real GDP, the lagged unemployment rate, a flag for COVID-19, and a trend variable. These mirror the controls used in my EFC regression in Table 1, minus any control variables that do not inform the range in EFC, such as year in college, because cross-admits should be new applicants to both Defendants. I also added the count of Defendants that admitted the student and the range of the rankings of the admitted schools, as these can inform the relative desirability of the institutions and the quality of the applicant, because an applicant who is admitted to more than two Defendants may be a stronger applicant than one who was only admitted to two Defendants.

TABLE 2: RANGE IN EFCs FOR CROSS-ADMITTED STUDENTS		
	Dependent Variable: <i>Real EFC Range</i>	
	(1)	(2)
Conduct	-2,954.98***	-1,718.24***
	(0.000)	(0.000)
Average Real Income		86.32***
		(0.000)
Average Real Net Worth		2.86
		(0.269)
Average Number in College		-5,037.25***
		(0.000)
Unemployment (1-year lag)		-219.73***
		(0.000)
Admitted Defendant Count		-773.79***
		(0.000)
Ranking Range		55.25*
		(0.055)
COVID		-2,003.30***
		(0.000)
Trend		580.00***
		(0.000)
Real GDP		-668.66*
		(0.072)
Observations	116,704	116,461
R-Squared	0.001	0.116

Notes: Robust p-values indicated as follows: ***p<0.01, **p<0.05, *p<0.1. Real GDP is measured in millions of dollars. Average income and average assets are measured in real dollars.

46. Table 2 shows that the absolute difference between the EFCs is lower when both Defendants for the cross-admitted student are in the 568 Group than when one or both Defendants are not. These results are economically and statistically significant for both specifications. I repeat this analysis using the Effective Institutional Price and find similar results.¹²⁷ The results from the Effective Institutional Price cross-admit regression are shown in Table 3 below. Together, these cross-admit regression results indicate that the EFC and the resulting Effective Institutional Price are closer together when the Challenged Conduct is in effect than when the Challenged Conduct is not. The cross-admit regression results demonstrate the chilling effect on competition of Defendants' participation in the 568 Group and contradict Defendants' Experts' ill-founded critique that the lack of identical EFC awards rules out anticompetitive effects.

127. In both the EFC cross-admit regression and Effective Institutional Price cross-admit regression, the R-squared values are low. Economists generally advise against putting too much emphasis on R-squared values. *See, e.g.,* WOOLDRIDGE at 39 (“Students who are first learning econometrics tend to put too much weight on the size of the R-squared in evaluating regression equations. For now, be aware that using R-squared as the main gauge of success for an econometric analysis can lead to trouble.”).

TABLE 3: RANGE IN EFFECTIVE INSTITUTIONAL PRICES FOR CROSS-ADMITTED STUDENTS

	Dependent Variable: <i>Real Effective Institutional Price Range</i>	
	(1)	(2)
Conduct	-6,711.40***	-4,735.54***
	(0.000)	(0.000)
Average Real Income		-28.66***
		(0.000)
Average Real Net Worth		-0.47
		(0.248)
Number in College		1,435.14***
		(0.000)
Unemployment (1-year lag)		263.09***
		(0.000)
Admitted Defendant Count		280.41***
		(0.000)
Ranking Range		423.58***
		(0.000)
COVID		500.38
		(0.123)
Trend		512.98***
		(0.000)
Real GDP		134.65
		(0.473)
Observations	110,491	110,275
R-Squared	0.020	0.089

Notes: Robust p-values indicated as follows: ***p<0.01, **p<0.05, *p<0.1. Real GDP is measured in millions of dollars. Average income and average assets are measured in real dollars.

B. Defendants’ Experts Misleadingly Argue That Defendants’ Actions Are Consistent with The Rest of The Higher Education Industry

47. In my Initial Report, I provided record evidence describing Defendants’ purpose in establishing the CM as being one to achieve “a consistent formula of need analysis.”¹²⁸ This evidence consisted of documents showing that Defendants did not want “competitive pressure”¹²⁹ and wanted to “avoid bidding wars.”¹³⁰ I also provided evidence that Defendants had engaged in the sharing of

128. Singer Report ¶126 (citing VANDERBILT-00000528 at -529 (statement from 568 Group Technical Committee Chair to 568 Group Members about renewing the 568 exemption)).

129. *Id.* (citing MITLIT-000078586 at -591).

130. *Id.* (citing DARTMOUTH_0000359527).

CSI, and that this information sharing is consistent with the economic criteria of monitoring output and prices, which economists view as indicators of a cartel.¹³¹

48. Defendants' Experts proffer various arguments that amount to claiming that Defendants would have behaved identically in the actual world compared to a but-for world, regardless of whether they had engaged in the Challenged Conduct, because all aspects of the Challenged Conduct are standard higher education industry practices.¹³² Dr. Stiroh and Dr. Long claim elements of the Challenged Conduct, including the application of applying professional judgement, the use of the IM as a starting point for the ultimate EFC calculation methodology, and the use of COFHE data for information-sharing are all typical practices for higher education institutions.¹³³ This line of argument is without merit.

49. There is a robust economic literature and theory describing how institutions act under competition versus in coordination. As explained in my Initial Report, economists describe the behavior of cartels as including the (1) monitoring output or prices; (2) developing organizations to effectuate cartel policies; and (3) developing inducements to support collusion, among other qualitative evidence of collusion.¹³⁴ I also detailed in my Initial Report the extensive record evidence indicating that the 568 Group exhibited these cartel-like behaviors, including the 568 Group's data exchange through COFHE, and regular surveys and meetings to monitor prices and maintain group policies and coordination.¹³⁵ Unilaterally, Defendants may have wanted to increase Effective Institutional Prices, but would not have done so in a competitive world because, without competitors

131. *Id.* ¶¶200-207.

132. Stiroh Report ¶¶83, ¶85. Long Report ¶¶215-220, ¶¶222-223, ¶¶229-230, ¶¶234-236, ¶¶240-244.

133. *See n. 131, supra.*

134. Singer Report ¶194.

135. *Id.* §II.D.2.

also raising prices, the price increase would induce some students to apply to (and potentially attend) a different school.¹³⁶

50. Dr. Stiroh recognizes that the unilateral interests of Defendants were to not get out of alignment with peer schools when it came to overall prices.¹³⁷ Under competition, Defendants would be incentivized to compete more vigorously for students than they would with the protection of knowing that their competitors will not offer additional aid. Competitive pressures would have caused schools to abandon financial aid principles that cause higher net prices than their competitors.

51. The record evidence bears out this point. For example, when Yale departed the 568 Group, then-Director of Student Financial Services, Caesar Storlazzi, stated “[b]y leaving the 568 Group, Yale is now free to give families more aid than they would have gotten under the consensus methodology.”¹³⁸ This suggests that, absent the collusion of the 568 Group, schools would provide additional aid as a form of competition for students. Dr. Stiroh also acknowledges that “even if the Challenged Conduct inflated EFCs . . . competitive forces would still discipline net prices through the other components of net price (e.g., packaging).”¹³⁹ While Dr. Stiroh is wrong about competition in the context of the Challenged Conduct, she correctly concedes that absent collusion Defendants would naturally discipline each other’s prices in a variety of ways, not only through EFCs. A cartel could stifle these methods of disciplining competitors’ net prices.

136. See, e.g., JEFFREY M. PERLOFF, MICROECONOMICS, 221 (Pearson 7th ed. 2015) (“Economists say that a market is competitive if each firm in the market is a price taker: a firm that cannot significantly affect the market price for its output or the prices at which it buys inputs.”); *id.* at 222 (“If buyers know that different firms are producing identical products and they know the prices charged by all firms, no single firm can unilaterally raise its price above the market equilibrium price. If it tried to do so, consumers would buy the identical product from another firm.”); N. GREGORY MANKIW, PRINCIPLES OF MICROECONOMICS, 293 (Cengage 8th ed. 2016) (“A competitive firm is small relative to the market in which it operates and, therefore, has no power to influence the price of its output. It takes the price as given by market conditions. By contrast, because a monopoly is the sole producer in its market, it can alter the price of its good by adjusting the quantity it supplies to the market.”).

137. Stiroh Report ¶102.

138. Caitlin Roman, *University Leaves Financial Aid Group*, YALE DAILY NEWS (Sept. 26, 2008), <https://yaledailynews.com/blog/2008/09/26/university-leaves-financial-aid-group/>.

139. Stiroh Report ¶140.

52. There is a wealth of qualitative evidence indicating that Defendants were coordinating in a manner that is consistent with the alleged conspiracy and is inconsistent with competition. As explained in my Initial Report, the 568 Group created and coordinated group requirements of the financial aid process through regular communications and the sharing of CSI.¹⁴⁰ This coordination was only allowed by the 568 exemption. Defendants knew that without the exemption, they would not be allowed to engage in the Challenged Conduct. This knowledge is why Defendants lobbied for the extension of the 568 exemption. Defendants believed that “[e]xpiration of the antitrust exemption would block the implementation of the consensus methodology.”¹⁴¹ If there was no difference between what Defendants and other Elite Private Universities used for financial aid principles, as Defendants’ Experts opine, then Defendants would never have needed to enter into the alleged 568 Agreement or receive an antitrust exemption to form the 568 Group.¹⁴² Rather, as the economics literature explains, the purpose of an agreement such as the 568 Agreement is that it provides an enforcement framework that allows a cartel to achieve supracompetitive prices.¹⁴³

53. Dr. Long attempts to refute my description of the information sharing component of the Challenged Conduct. She claims that COFHE data are widely available and used by both Defendants and non-Defendants. This is a misleading statement. The COFHE colorbooks and online database were not widely available to universities outside of the 568 Group. There are only 39

140. Singer Report §II.D.

141. DUKE568_0155589 at -613. *See also* Columbia_00058983 (President DeGioia remarking that 568 exemption was “fundamental to the group’s ability to . . . agree upon a need-based financial aid system.”).

142. *See, e.g., Memorandum of Understanding, The 568 Presidents’ Group*, <https://web.archive.org/web/20050311231218/http://568group.org/docs/memo.pdf> (archived 11 Mar. 2005) [hereafter *Memorandum of Understanding*] (under the “Purposes of the group” is listed: “2. Discuss and agree upon common principles of analysis for determining the financial need of undergraduate financial aid applicants.”).

143. Margaret C. Levenstein and Valerie Y. Suslow, *What Determines Cartel Success?*, 44(1) JOURNAL OF ECONOMIC LITERATURE 43–95, 44 (2006) (“Following Stigler (1964), many economists have argued that such attempts would inevitably fail as colluding firms succumbed to the temptation to shave prices secretly in order to increase their individual firm profits. Subsequent theoretical work has established that sufficiently patient and well-informed firms can, in principle, collectively deter cheating and allow collusion to survive.”).

universities that are COFHE members, including 16 of the 17 Defendant institutions.¹⁴⁴ While there is overlap with financial aid information compiled by COFHE and that reported by IPEDS, the COFHE data are disseminated to COFHE members via the “colorbooks” with a one-year lag versus a three-year lag with IPEDS.¹⁴⁵ Although there is not perfect overlap between the COFHE group and the 568 Group, the 568 Group also received non-publicly available data through surveys, meetings, and other group communications, as detailed in my Initial Report.¹⁴⁶ These alternative methods of exchanging CSI could be used to coordinate cartel activities.

54. Dr. Long also claims that I “ignore the importance of data in running a higher education institution” due to my statement that the 568 Group shared CSI.¹⁴⁷ While I agree with Dr. Long that benchmarking data are widely used and relevant to higher education, her arguments ignore the reality of the preferential and coordinated data the Defendants received, which deviates from the general benchmarking data used by other universities. For example, Vanderbilt put together a June 2016 presentation to analyze “competitor financial aid data” of 2015 matriculants using COFHE financial aid data.¹⁴⁸ This presentation included the average grant-aided parent contribution, median grant-aided parent income, average grant, and average net price for Vanderbilt and 13 competitors, including nine Defendants.¹⁴⁹ The presentation also included total income, net worth, and total parent contribution (broken down as the portion from income and assets, and showing the contribution as a percentage of income).¹⁵⁰ Importantly, these data were not widely available to universities outside of

144. These 39 universities and colleges include my relevant market, minus Carnegie Mellon and Notre Dame. *Consortium on Financing Higher Education*, MIT, <https://web.mit.edu/cofhe/> (last visited Oct. 6, 2024).

145. For example, the latest IPEDS financial aid data currently available is for 2021-22.

146. Singer Report §II.D.2. COFHE data exchanges are not inherently legal either, meaning that just because a small number of non-Defendants have access to COFHE data does not indicate that COFHE data are also not competitively sensitive data.

147. Long Report ¶240.

148. VANDERBILT-00285876.

149. *Id.* at -877-78 (The nine Defendants listed as competitors are Yale, Chicago, Dartmouth, Columbia, Duke, Penn, Rice, Brown, and Northwestern.).

150. *Id.* at -882.

the 568 Group. The recency of information such as the average net price by income group helps facilitate price-setting decisions.¹⁵¹ This and other data used by the Defendants (such as from the 568 surveys and meetings) extend beyond the general benchmarking data available to non-Defendants.

55. Finally, Dr. Stiroh makes a number of comparisons between the CM guidelines to the Base IM calculations, arguing that the differences between the two methodologies indicates that the CM was potentially beneficial to certain Class Members.¹⁵² It is important to note, however, that the CM influenced the IM in addition to the IM being a starting point for the CM. Defendants worked with the College Board both by having 568 Group members serve on the College Board and its relevant committees to make changes to the Base IM and by recommending specific changes to the Base IM.¹⁵³ Cornell even stated in a 2022 email that “it is hard to define what is 568 vs what is IM. In some cases, the 568 recommendation is to use IM, in other cases the 568 recommendation became IM.”¹⁵⁴ Given this significant overlap and cross-influence between the IM and the CM, Dr. Stiroh’s comparison are merely comparing a formula impacted by the 568 Group with another formula impacted by the 568 Group.

56. Dr. Stiroh’s argument also ignores my quantitative findings. My EFC overcharge regression illustrates that the Challenged Conduct is associated with an increase in EFCs.¹⁵⁵ This indicates that even if the CM EFC was lower than the Base IM EFC for some students, the CM EFC would likely still be artificially inflated by the Challenged Conduct. In my Initial Report, I showed via in-sample prediction that all or nearly all Class Members paid higher Effective Institutional Prices

151. See, e.g., Margaret C. Levenstein and Valerie Y. Suslow, *Cartel bargaining and monitoring: The role of information sharing*, in *The Pros and Cons of Information Sharing*, SWEDISH COMPETITION AUTHORITY (2006) at 25 (“Cartels often go to great efforts to increase the frequency of reporting, suggesting that they believe that the increase in communication will prevent cheating and facilitate collusion.”).

152. Stiroh Report ¶¶115-122.

153. Singer Report ¶159.

154. CORNELL_LIT0000112401.

155. See Table 1.

as a result of the Challenged Conduct.¹⁵⁶ Therefore, even if Dr. Stiroh were able to find a specific instance in which the CM resulted in a more generous EFC than the Base IM, the Class Member would still be harmed – either because the Class Member’s CM EFC was still inflated relative to the but-for world EFC, and/or due to the Class Member’s financial aid packaging being directly affected by the Challenged Conduct.

III. COMMON IMPACT

57. In Part III of my Initial Report, I employed a standard, two-step approach to demonstrate common impact both of which steps are common to the Class as a whole. In the first step, I estimated a regression model and found that the Challenged Conduct caused Effective Institutional Prices at Defendant institutions to increase above the levels that would have prevailed in its absence (that is, “but-for” the Challenged Conduct). I also cited other qualitative evidence that the Challenged Conduct would lead to artificially inflated Effective Institutional Prices, including economic theory and documentary evidence.¹⁵⁷ In the second step, I demonstrated with classmate

156. Singer Report §III.B.1.

157. See Singer Report Part III.A. See also, e.g., GTWNU_0000317271 at -272 in the first admissions cycle after implementing the CM (the fall of 2003), President Schapiro, then Chair of the 568 Group, wrote “awards offered to common admits were, for the first time in many years, reasonably consistent across participating institutions); VANDERBILT-00047310 (“Vanderbilt benefits from this exemption, having the opportunity to discuss the methodology used to determine eligibility for need-based financial aid with similar institutions. This discussion allows for a common approach, so a family’s expected family contribution does not vary to any great extent from school to school. *It helps to avoid bidding wars between schools*, so families can focus on other aspects of selecting on other aspects of selecting an institution; not focusing on the lowest net price.”); Caitlin Roman, *University Leaves Financial Aid Group*, YALE DAILY NEWS (Sept. 26, 2008), <https://yaledailynews.com/blog/2008/09/26/university-leaves-financial-aid-group/>. (“By leaving the 568 Group, Yale is now free to give families more aid than they would have gotten under the consensus methodology.”); *Yale Cuts Costs for Families and Students*, YALENEWS (Jan. 14, 2008), <https://news.yale.edu/2008/01/14/yale-cuts-costs-families-and-students> (With its announcement to leave the 568 Group, Yale adopted an income threshold of \$60,000, under which there would be no parental contribution to the COA); GTWNU_0000009789 at -791 (in December 2008, Mr. Belvin told the 568 Group that “[i]nstitutional policy statements should not modify the aid office’s requirement to evaluate all applications using the Consensus Approach,” stating that waiving tuition for families with incomes below \$60,000 should “not . . . be . . . automatic”); UCHICAGO-0000052966 at -967 (The University of Chicago acknowledged in 2013 that “568 Group – it is hampering our ability to compete.”); Emory_568Lit_0006214 (in purportedly leaving the 568 Group, in 2012, Emory officials lamented that the Group “generates a restrictive environment for packaging financial aid at a time when college and universities need to be more flexible and responsive”); PENN568-LIT-00133526 (“Harvard and Princeton never joined because they did not want to limit the usage of their funds.”); Columbia_00056363 at -364 (noting that Harvard, Yale, and Princeton could be “significantly more generous” and are not applying the CM).

evidence and analyses that this generalized artificial overcharge in Effective Institutional Prices would have transmitted to all or nearly all Class Members, thereby producing classwide impact. Below, I expound Defendants' Experts' responses to this evidence, and I provide my rejoinders and/or concessions to their claims. I find that none of Defendants' Experts' responses fundamentally alters my opinions that the Challenged Conduct resulted in an artificial inflation of Effective Institutional Prices paid by Class Members, and that this artificial inflation impacted all or nearly all Class Members.

A. Classwide Evidence Is Consistent with the Challenged Conduct Causing an Artificial Inflation in Effective Institutional Prices for All Defendants

1. Qualitative Evidence Supports the Claim that the Challenged Conduct Would Have Resulted in Artificially Inflated Effective Institutional Prices

58. In Part III.A.1 of my Initial Report, I provided record evidence that was consistent with the Challenged Conduct having artificially inflated Effective Institutional Prices. This evidence supported the assertion that the 568 Group reduced competition across Defendants, inflating Effective Institutional Prices.¹⁵⁸ For instance, Vanderbilt stated that the 568 exemption helped to “avoid bidding wars between schools,”¹⁵⁹ and Yale stated that it left the 568 Group in 2008 to allow itself to offer more generous aid as a means to compete for students.¹⁶⁰ I also provided record evidence from Harvard, Stanford, and Princeton that showed that these institutions did not join the 568 Group because doing so would limit their ability to offer more generous institutional grant aid.¹⁶¹ When viewed through an economic lens, evidence of competitors consistently meeting to discuss how they

158. Singer Report §III.A.1.

159. *Id.* ¶221. *See also* McWade Dep. 190:3-191:4 (568 members “wanted to see if we could agree on how to analyze families’ ability to pay,” and were concerned that “after the exemption expired, oh, we’re going back to the Wild Wild West where schools would do . . . what they wanted and different things.”).

160. Singer Report ¶222.

161. *Id.* ¶226.

will go about setting Effective Institutional Prices in the future, and to share information that is not publicly available about those prices, is consistent with a dampening of competition.¹⁶²

59. Dr. Stiroh asserts that “there is no economic nexus between an alleged agreement on certain principles, starting points, and approaches for assessing family ability to pay and the ultimate amount of aid any student was offered, or the net price charged by any Defendant school to any proposed Class member.”¹⁶³ Dr. Stiroh is wrong and is contradicted by the qualitative and quantitative evidence described in my Initial Report and summarized above. Her argument also ignores the multiple avenues through which the Challenged Conduct could have artificially inflated Effective Institutional Prices.

60. The calculation of EFCs and the structure of packaging would bear on the Effective Institutional Prices. As described above, the Challenged Conduct artificially inflated EFCs, which then translated into higher Effective Institutional Prices. Qualitative evidence is consistent with the Challenged Conduct, via the affordability principle, also altering packaging, resulting in Defendants substituting away from institutional gift aid to loans.¹⁶⁴ Dr. Stiroh argues that the Challenged Conduct did not artificially inflated EFCs and that higher EFCs would not necessarily result in higher Effective Institutional Prices. Dr. Stiroh dismisses my claim that the Challenged Conduct had affected packaging, instead claiming that packaging was “not alleged to have been affected by the agreement,” ignoring the qualitative evidence provided in my Initial Report suggesting otherwise.¹⁶⁵

162. *Id.* ¶226. *See also* McWade Dep. 138:22-139:19 (testimony by former Head of 568 Group Technical Committee that Harvard and Princeton never joined the 568 Group because “they wanted to be more generous in the analysis of the family’s ability to pay”).

163. Stiroh Report ¶70.

164. Singer Report ¶226.

165. Stiroh Report ¶114. In §II.A.4 of my Initial Report, I explained how packaging would have been affected by the Challenged Conduct.

61. Dr. Stiroh asserts that the Challenged Conduct did not result in an artificial increase in EFCs among all Class Members.¹⁶⁶ As support for her argument, she compares Defendants' students' actual EFCs to their EFCs under alternative methodologies.¹⁶⁷ Dr. Stiroh first compares EFCs for students that were cross admitted by both Defendants and the other five universities included in the Elite Private University services market ("non-Defendants"), and she finds that many non-Defendants set higher EFCs than Defendants.¹⁶⁸ She then compares EFCs for cross-admits to both participating and non-participating Defendants in the same academic year.¹⁶⁹ She claims that there is no clear uniform pattern to show that for non-participating Defendants offered lower EFCs than participating Defendants.¹⁷⁰ Lastly, Dr. Stiroh compares students' actual EFCs to their Base IM EFCs for the three Defendants that produced data containing both actual and Base IM EFCs—Cornell, Duke, and Vanderbilt.¹⁷¹ She finds that 34 percent of students analyzed had a higher EFC under the Base IM.¹⁷²

62. Dr. Stiroh's EFC comparison analyses are flawed because her benchmark (IM) EFCs would have also been contaminated by the Challenged Conduct. I noted this point in my Initial Report when I explained that the Challenged Conduct would have been expected to produce an "umbrella effect," which is a common economic phenomenon in price-fixing conspiracies.¹⁷³ Given Defendants' significant collective prestige and market power, EFCs for other Elite Private Universities and for non-participating Defendants would likely be at least somewhat inflated by the Challenged Conduct, rendering those EFCs uninformative as to what EFCs would have looked like

166. Stiroh Report ¶113.

167. *Id.*

168. *Id.* ¶¶131-132. These five non-Defendant universities are Carnegie Mellon, Harvard, Princeton, Stanford, and Washington University of St. Louis.

169. *Id.* ¶¶134-135.

170. *Id.* ¶135.

171. *Id.* ¶¶136-137.

172. *Id.* ¶137.

173. Singer Report ¶228.

in a but-for world. In fact, in my Initial Report, I cited qualitative evidence explaining that the Challenged Conduct likely would have altered the Base IM through (a) Defendants making recommendations about and suggesting changes to the Base IM itself, and (b) having Group members serve on the College Board and its relevant committees tasked with making changes to the Base IM.¹⁷⁴ Hence, Base IM EFCs would not be representative of what EFCs would have been absent the Challenged Conduct, and Dr. Stiroh's comparison of Defendants' EFCs to Base IM EFCs is therefore invalid.

63. Dr. Stiroh states that "Dr. Singer's economic modeling cannot provide a prediction of the expected changes that any individual school would have made to its EFC calculations absent the Challenged Conduct."¹⁷⁵ In Part II.A above, I respond to Dr. Stiroh's claim by empirically testing what Defendants' EFCs would have looked like absent the Challenged Conduct. Dr. Stiroh's hypothesis is that the Challenged Conduct did not result in an artificial inflation in EFCs. I reject her hypothesis—Defendants' EFCs were artificially inflated by approximately \$515 per Class Member-academic year, and this result is both economically and statistically significant. This quantitative evidence is consistent with the expected economic effect of the Challenged Conduct on EFCs based on the qualitative evidence, and it is inconsistent with Dr. Stiroh's claim that EFCs were not affected by the Challenged Conduct.

64. Dr. Stiroh asserts that there is no clear economic linkage between higher EFCs and higher Effective Institutional Prices, because Defendants would have competed on packaging.¹⁷⁶ Put

174. *Id.* ¶159. PENN568-LIT-00029001 ("My recollection is that the tables were developed by a FASSAC subcommittee which included some 568 cross-representation."); AMHE-00012771 at -783 ("Strong feeders for FASSAC have been identified as: . . . The 568 Group."); COFHE-02-00011341 at -343 ("The GAO also discovered that many of the 568 Group recommendations for methodological revisions that better serve students and parents had been adopted by the College Board as well as by individual institutions.").

175. Stiroh Report ¶109.

176. *Id.* ¶¶138-143.

differently, her argument suggests that competition would have resulted in Defendants substituting away from loans and/or work-study awards to greater family contributions. She bases this on the fact that several Defendants have enacted no-loan policies and/or varying work-study policies over time.¹⁷⁷

65. Dr. Stiroh's argument suffers from a timing problem. That is, her argument does not correspond to the time when Defendants enacted no-loan policies. For instance, Yale, Rice, and Chicago all adopted no-loan policies, but they all did so *immediately after* having ended their participation in the Challenged Conduct.¹⁷⁸ Similarly, Brown and Emory adopted no-loan policies but did so multiple years after having left the 568 Group.¹⁷⁹ The evidence for these Defendants contradicts Dr. Stiroh's argument—Defendants only began to compete on packaging *after* they had ceased to participate in the Challenged Conduct—consistent with packaging also having been impacted by the Challenged Conduct.

66. Dr. Stiroh's argument also contradicts economic sense. Her premise assumes that Defendants would have agreed to a common formula that increased EFCs, but then simply counteracted these higher EFCs by reducing loans or work-study. Dr. Stiroh provides no justification for why Defendants would have agreed merely to substitute from self-help aid (which requires a student to pay back any loans or meet additional financial need through paid labor) to out-of-pocket payments.

177. Stiroh Report ¶¶140-141.

178. Jon Victor, *Despite Perkins expiration, little impact on campus*, YALE NEWS, (Oct. 9, 2015), <https://yaledailynews.com/blog/2015/10/09/despite-perkins-expiration-little-impact-on-campus>; Long Report Figure 9 (showing Chicago enacted a no-loan policy in the 2015 academic year. My Class Period definition counts Chicago as leaving the 568 Group starting in the 2015 academic year); Stiroh Report ¶140 (stating that Rice no longer had institutional loans starting in 2022. My Class Period definition counts Rice as leaving the 568 Group starting in 2022.).

179. Natalie Villacres, *Brown makes elimination of undergraduate loans permanent*, BROWN DAILY HERALD, (Mar. 23, 2023), <https://www.browndailyherald.com/article/2023/03/brown-makes-elimination-of-undergraduate-loans-permanent>. Stiroh Report ¶140 (stating that Emory began a no-loan policy in 2022. My Class Period definition counts Emory as having left the 568 Group starting in 2013).

67. Dr. Stiroh's argument is also at odds with the evidence of Defendants' fear of "bidding wars" described above. And Dr. Stiroh's argument is inconsistent with an economic rationale for either the existence of the 568 Group or for the 568 exemption itself.

68. Furthermore, Dr. Stiroh's hypothesis that higher EFCs do not result in higher Effective Institutional Prices can be directly tested. In Table 4, I run regressions of Effective Institutional Price on EFCs to measure how Effective Institutional Price increases in response to a one-dollar increase in EFC. Based on my primary regression specification in column 6, I find that every one-dollar increase in real EFC corresponds to a \$0.24 increase in real Effective Institutional Prices while holding all other factors constant. This result is both economically and statistically significant. These results are inconsistent with Dr. Stiroh's assertion that higher EFCs do not necessarily translate into higher Effective Institutional Prices.

TABLE 4: REGRESSIONS OF REAL EFFECTIVE INSTITUTIONAL PRICE ON REAL EFC

	Dependent Variable: <i>Real Effective Institutional Price</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
	<i>No Student Fixed Effects</i>			<i>Includes Student Fixed Effects</i>		
Real EFC	0.33*** (0.000)	0.33*** (0.000)	0.33*** (0.000)	0.24*** (0.000)	0.24*** (0.000)	0.24*** (0.000)
Adj. Gross Income	26.21*** (0.000)	26.20*** (0.000)	26.25*** (0.000)	18.24*** (0.000)	18.12*** (0.000)	17.94*** (0.000)
Net Worth	0.64*** (0.007)	0.63*** (0.007)	0.63*** (0.007)	0.11** (0.018)	0.10** (0.022)	0.09** (0.032)
Number in College	-1,444.06*** (0.000)	-1,436.66*** (0.000)	-1,444.78*** (0.000)	-4,146.24*** (0.000)	-4,131.06*** (0.000)	-4,120.30*** (0.000)
Year in College	1,043.02*** (0.000)	1,032.76*** (0.000)	1,064.22*** (0.000)	1,537.42*** (0.000)	1,389.45*** (0.000)	-1,110.64*** (0.000)
Student's Gift Aid from Other Sources	0.12** (0.017)	0.12** (0.015)	0.12** (0.018)	0.35*** (0.000)	0.35*** (0.000)	0.35*** (0.000)
Lagged Excess Endowment Investment Returns		-265.93 (0.130)	104.01 (0.512)		184.58 (0.264)	105.47 (0.532)
Inst. Tuition Rev per FTE Undergraduate (1-year lag)		0.02*** (0.004)	0.02*** (0.000)		0.10*** (0.000)	0.04*** (0.000)
% of FY-FTE Undergrads Receiving Financial Aid		-74.25*** (0.000)	-72.55*** (0.000)		-10.15*** (0.005)	-22.72*** (0.000)
Unemployment (1-year lag)			-141.53*** (0.000)			-108.75*** (0.000)
COVID			4,339.66*** (0.000)			3,513.54*** (0.000)
Trend			-187.71*** (0.000)			2,317.76*** (0.000)
Real GDP			0.27*** (0.001)			0.58*** (0.000)
Observations	690,236	690,236	690,236	690,236	690,236	690,236
R-Squared	0.37	0.37	0.37	0.88	0.88	0.88
Includes Institution Fixed Effects?	Y	Y	Y	Y	Y	Y
Includes Institution*Student Fixed Effects?	N	N	N	Y	Y	Y
Number of Fixed Effects	17	17	17	256,661	256,661	256,661

Notes: Robust p-values in parentheses; ***p<0.01, **p<0.05, *p<0.1. Adjusted Gross Income and Net Worth reported in thousands. All dollar denominated variables are in real dollars. I use regression data that is adjusted based on the adjustments described in Part III.A.c.

69. Dr. Stiroh also outright dismisses the qualitative evidence set forth in my Initial Report showing that Challenged Conduct would have affected packaging, and instead she states that packaging is “not alleged to have been affected by the agreement.”¹⁸⁰ I showed that qualitative evidence is consistent with Defendants having applied the affordability principle to packaging, and this would serve as another avenue through which the Challenged Conduct could impact Effective

180. Stiroh Report ¶114. In §II.A.4 of my Initial Report, I explained how packaging would have been affected by the Challenged Conduct.

Institutional Prices besides through its effect on EFCs.¹⁸¹ A key premise of the Overarching Agreement was that “[i]t is quite reasonable for parents and students to borrow to pay for college.”¹⁸²

70. Applying Dr. Stiroh’s own logic to my econometric findings necessarily implies that Defendants would have colluded on more than just EFCs. As alluded to earlier in Part II, Dr. Stiroh argues that, even if Defendants had artificially inflated EFCs, “competitive forces would still discipline net prices through the other components of net price (e.g., packaging).”¹⁸³ By Dr. Stiroh’s own reasoning, if it were the case that Effective Institutional Prices were at competitive levels even while EFCs were subject to collusion, then my Effective Institutional Regressions would show a zero overcharge. That my Effective Institutional Price regressions show an overcharge thereby implies that Effective Institutional Prices were not at competitive levels.

2. Defendants’ Experts Critiques of My Overcharge Model Are Without Merit

71. The first step of my two-step approach to demonstrating common impact is showing that Defendants’ participation in the Challenged Conduct caused a generalized artificial overcharge in Effective Institutional Prices. In addition to pointing to qualitative evidence linking the Challenged Conduct with artificially inflated prices through an economic lens, I showed this quantitatively using a “before-after” multiple regression methodology, which is a standard empirical approach widely applied in both research and litigation.¹⁸⁴ My methodology consists of regressing Effective Institutional Prices on a conduct dummy variable, which equals to one during periods when a Defendant had participated in the Challenged Conduct, and is equal to zero during clean periods when

181. Singer Report §II.A.4.

182. Sandy Baum, A PRIMER ON ECONOMICS FOR FINANCIAL AID PROFESSIONALS, COLLEGE BOARD AND THE NATIONAL ASSOCIATION OF STUDENT FINANCIAL AID ADMINISTRATORS 4 (1996) at 54. *See also* Singer Report ¶34.

183. Stiroh Report ¶140.

184. Theon van Dijk & Frank Verboven, *Quantification of Damages*, 3 ISSUES IN COMPETITION LAW AND POLICY 2331-2348, 2331-2332 (ABA Section of Antitrust Law 2008) (“The concept underlying most economic damages assessments is that of the ‘but-for’ world.”).

the Defendant had not participated. I included numerous student-specific, Defendant-specific, and macroeconomic control variables which allow me to isolate the effect of the Challenged Conduct on Effective Institutional Prices. I also showed results using both Defendant and Defendant-student fixed effects, which control for time-invariant unobservable factors that may influence Effective Institutional Prices. In my primary regression model, which includes the largest number of control variables and uses Defendant-student fixed effects, I obtained a conduct coefficient of 1,497. This result implies that a Defendant's participation in the Challenged Conduct caused their Effective Institutional Prices to be artificially inflated by \$1,497 per Class Member and academic year while holding all other factors constant.¹⁸⁵

72. Dr. Hill and Dr. Stiroh both critique aspects of my regression implementation. For instance, Dr. Hill critiques my use of robust standard errors, data-processing and control-variable construction, and which control variables I consider. Dr. Stiroh critiques my use of a common overcharge and the benchmark periods I treat as "clean" periods. Even after I accept a material number of Dr. Hill's data-processing adjustments, I continue to find an economically and statistically significant artificial overcharge of \$1,202 per Class Member and academic year.¹⁸⁶

73. My overcharge estimates are conservative and therefore understate the actual artificial overcharge paid by Class Members for multiple reasons. First and foremost, as demonstrated in Part I of my Initial Report, Defendants collectively held market power in the Elite Private University Market during the Class Period. Therefore, the Challenged Conduct would be expected to artificially inflate Effective Institutional Prices throughout the Elite Private University Market due to what economists refer to as the umbrella effect. Second, my regressions account for differences in Effective Institutional Prices arising from periods when Defendants participated and did not participate in the

185. Singer Report ¶249.

186. See Part III.A.2.c.

Challenged Conduct. Record evidence suggests that Defendants that ceased to participate often did not fully withdraw from the 568 Group, thereby suggesting that their Effective Institutional Prices were likely still artificially inflated during non-participation periods. For instance, many Defendants had still engaged in the information-exchange element of the Challenged Conduct even during non-participatory years.¹⁸⁷ Third, economists have found that there is often a lag in timing between when a cartel ends and when market prices return to competitive levels, and this price reduction lag is greater the longer the cartel was in place.¹⁸⁸ Given that the Challenged Conduct spans nearly two decades, one would expect that post-2022 prices are likely still at least partially inflated as a result of the alleged conspiracy—thereby rendering my overcharge calculations conservative.

a. Dr. Stiroh's Critique of My Use of a Common Overcharge Is Unfounded

74. To measure the effect of the Challenged Conduct on Effective Institutional Prices, I use a standard approach that involves regressing real Effective Institutional Prices on a dummy variable that captures when a Defendant had or had not participated in the Challenged Conduct. Because my regressions use a single conduct variable across all Defendants and academic years, these regressions each produce a single, common overcharge. Dr. Stiroh critiques my use of a common overcharge, suggesting that I should instead estimate Defendant-specific and time-period-specific overcharges. Below, I describe the numerous flaws in her arguments.

75. The fundamental issue with Dr. Stiroh's argument is that one should use all of Defendants' reliable data in an econometric model for the specific purpose of providing a reliable

187. There were two primary sources of information exchange. One involved the 568 surveys, meetings, and communications. The other was the COFHE meetings. Many Defendants who came in and out of 568 remained in COFHE the entire time. Singer Report ¶¶165-166.

188. Joseph E. Harrington, Jr., *Post-Cartel Pricing During Litigation*, 52(4) THE JOURNAL OF INDUSTRIAL ECONOMICS 517-533, 517 (2004) ("Standard methods in the U.S. for calculating antitrust damages in price-fixing cases are shown to create a strategic incentive for firms to price above the non-collusive price after the cartel has been dissolved. This results in an overestimate of the but for price and an underestimate of the level of damages. The extent of this upward bias in the but for price is greater, the longer the cartel was in place and the more concentrated the industry.").

measure of the overcharges, if any, attributable to the Challenged Conduct. The price effects of supply and demand factors influencing Effective Institutional Prices in the market are most reliably measured by a model across as much of the market as possible. Each student enrollment-payment transaction in each academic year plays some role in reliably measuring market-price effects for every other transaction. Dr. Stiroh attempts to estimate my regressions separately by Defendant, but this fails to provide reliable estimates of market-price effects, and therefore of the separate price effects of the Challenged Conduct, for the simple reason that she excludes significant portions of the available, relevant data from other Defendants. Dr. Hill obtains similar unsound, economically nonsensical results when he dismisses post-Class Period data, as I show in Part III.A.2.e. By limiting the data to narrow slices, Defendants' Experts exclude essential information that informs market effects on Effective Institutional Prices.

76. Similarly, my model provides the most reliable estimate of overcharges by including a single indicator for the Challenged Conduct across the whole Class Period. Defendants' Experts' choices to chop up the Challenged Conduct indicator into time segments, as Dr. Stiroh does, or to artificially mute any effect from the Challenged Conduct by including highly collinear dummy variables, as Dr. Hill does, increases the likelihood that market and conspiracy price effects are conflated, rendering their various estimates unreliable and unsound.¹⁸⁹

189. See Part III.A.2.d. Dr. Hill tests (1) incorporating time fixed effects, and (2) separate dummy variables for each year after the onset of COVID-19 into my regressions. Hill Report §8.3. Both of these erroneous adjustments would be expected to artificially dampen any effect of the Challenged Conduct on Effective Institutional Prices. Many Defendants provided data that only contain one year to two years in which the Defendant did not participate in the conspiracy (i.e., 2023 and 2024). Because the data I use are annual, a dummy variable for 2023 would absorb most of the conduct effect due to its high collinearity with the conduct variable, which is what Dr. Hill does when he adds time fixed effects or COVID-19 dummy variables for every year after the onset of COVID-19.

i. My Use of a Common Overcharge Across Defendants is Standard Practice

77. Dr. Stiroh opines that my use of a common overcharge is improper because it masks variation that may have occurred across Defendants.¹⁹⁰ For example, Dr. Stiroh argues that Defendants did not uniformly adopt all elements of the CM, and that many of the CM guidelines only had applied to certain groups of students.¹⁹¹ Dr. Stiroh claims that, as a result of this asymmetry, it is “economically reasonable to infer that impact, if any, would not be uniform across schools.”¹⁹² To remedy this purported issue, Dr. Stiroh runs the same regressions as those I produced in Table 11 of my Initial Report, but separately by Defendant.¹⁹³ Because she applies my regressions to each separate Defendant dataset, Dr. Stiroh obtains an individual conduct coefficient for each Defendant, rather than a common conduct coefficient for all Defendants. She also generates Defendant-specific coefficients on each of the control variables, without any *a priori* justification for doing so. Her results show that not all Defendants’ regressions report both a positive and statistically significant overcharge, and she therefore claims that these regressions do not demonstrate harm to all Class Members.¹⁹⁴

78. Before explaining the flaws in her methodology, it first bears noting that Dr. Stiroh’s Defendant-specific regressions using my primary regression model find positive and economically significant conduct coefficients for 12 out of the 15 Defendants to which she is able to apply my primary (student fixed effects) regression model.¹⁹⁵ For the two Defendants that Dr. Stiroh is unable to estimate using my primary regression model—Emory and Dartmouth, which only produced data

190. Stiroh Report ¶176.

191. *Id.* ¶172.

192. *Id.*

193. *Id.* §VII.A.1.

194. *Id.* ¶¶178-179.

195. *Id.* at Figure 7.3, Column (d).

pertaining to first years—my primary model without student fixed effects also finds positive and economically significant coefficients.¹⁹⁶ Thus, Dr. Stiroh’s Defendant-specific version of my primary regression specification reports a positive conduct coefficient for 14 out of 17 Defendants.¹⁹⁷ Similarly, my log specifications of the same model shows positive conduct coefficients for 15 out of 17 Defendants.¹⁹⁸ Dr. Stiroh’s application of my regressions separately by Defendant is flawed for numerous reasons.

79. To begin, Dr. Stiroh’s methodology ignores the scope of the regression data. As noted in my Initial Report, most Defendants had produced limited post-conduct data, and nine out of 17 Defendants did not produce data prior to the introduction of the alleged Challenged Conduct in 2003.¹⁹⁹ For these nine Defendants, Dr. Stiroh’s Defendant-specific regressions remove significant pre-conduct variation in Effective Institutional Prices that inform what Effective Institutional Prices would have looked like in a world but-for the Challenged Conduct. In my Initial Report, I explained that Effective Institutional Prices would have likely still been at least partially inflated by the existence of the alleged conspiracy once it had begun, even for those Defendants that may have left

196. *Id.* at Figure 7.2, Column (d). Dr. Stiroh’s regressions that include student fixed effects show values of “N/A” for Emory and Dartmouth. This is because my student-fixed-effects regressions require that there be more than one observation per student, as these regressions use within-student variation for identification. Because Emory and Dartmouth only produced first-year data, these Defendants therefore do not have the student-level variation required to estimate a student fixed effects model.

197. The three Defendants that Dr. Stiroh reports as having zero or negative conduct coefficient values in my primary levels regression are Caltech, Columbia, and Vanderbilt. *See Id.* at Figure 7.3, Column (d).

198. *Id.* at Figures 7.2-7.3, Column (d). Similar to my review of Dr. Stiroh’s levels results, I refer to the conduct coefficients from the student fixed effects regressions for all Defendants other than Emory and Dartmouth, and I refer to the school fixed effects results for Emory and Dartmouth. The two Defendants that Dr. Stiroh reports as having zero or negative conduct coefficient values in my log-linear regression are Caltech and Vanderbilt.

199. Singer Report ¶231; *id.* at Appendix 3 Table 1.

the 568 Group, as a result of the “umbrella effect.”²⁰⁰ Therefore, Dr. Stiroh’s Defendant-specific regression results likely produce downwardly biased conduct coefficients relative to my methodology, since the latter includes pre-Challenged Conduct data and uses these data as a comparison benchmark period for *all* Defendants during their participation periods.

80. Additionally, for five Defendants—Cornell, Georgetown, MIT, Notre Dame, and Rice—their data start during the Class Period and only contain a single academic year in which they were not subject to the Challenged Conduct (in 2023).²⁰¹ When Dr. Stiroh runs her Defendant-specific regressions for these five Defendants, she assumes that a single year can serve as a proxy for the benchmark period. This assumption is unrealistic. It assumes that any effects of the Challenged Conduct would have dissipated immediately. As explained in my Initial Report, economic theory suggests that post-conspiracy prices tend to remain at least partially inflated for some period of time after the dissolution of a price-fixing cartel.²⁰² Therefore, Dr. Stiroh’s Defendant-specific conduct coefficients for those five Defendants are subject to further downward bias due to any lingering price effects that may have contaminated their single clean year.

200. *Id.* ¶228. *See, e.g.,* Johannes Odenkirchen, *Pricing Behavior in Partial Cartels*, DÜSSELDORF INSTITUTE FOR COMPETITION ECONOMICS, No. 299 Discussion Paper (Sept. 2018), 2 (“Supporting standard theory, we find that a partial cartel is sufficient to distort market prices. Average market prices are higher when partial cartels form than without any cartel in the market. This confirms the expected umbrella effect.”). *See also* Roger D. Blair & Virginia G. Maurer, *Umbrella Pricing and Antitrust Standing: An Economic Analysis*, 4 UTAH LAW REVIEW, 763-796, 764 (1982) (“If the dominant firms fix prices, purchasers from the competitive fringe firms will still pay a price that exceeds what the market price would be in the absence of collusion. That result is mandated by the competitive fringe firms’ role as price takers. In other words, fringe firms will not act as though their output decisions have a perceptible influence on price. Accordingly, they charge the current market price and simply adjust their output level to maximize profits. Thus, the fringe firms set their prices under the ‘umbrella’ of the dominant firms.”).

201. I refer to the updated regression data that is described in Part III.A.2.c.

202. Singer Report ¶228. *See, e.g.,* Joseph E. Harrington, Jr., *Post-Cartel Pricing During Litigation*, 52(4) THE JOURNAL OF INDUSTRIAL ECONOMICS 517-533, 517 (2004) (“Standard methods in the U.S. for calculating antitrust damages in price-fixing cases are shown to create a strategic incentive for firms to price above the non-collusive price after the cartel has been dissolved. This results in an overestimate of the but for price and an underestimate of the level of damages.”). *See also* H. Peter Boswijk, Maurice J. G. Bun, and Maarten Pieter Schinkel, *Cartel Dating*, 34 JOURNAL OF APPLIED ECONOMETRICS 26-42 (2017) (showing that the comparison of before-after model (comparing actual to but-for prices) leads to an underestimation of overcharges when the begin and end dates of a cartel are not precisely dated).

81. Dr. Stiroh's implementation of my regressions separately by Defendant also alters the methodology employed by my model, and therefore produces inconsistent results to my own. My model measures the price effect of the Challenged Conduct by comparing Effective Institutional Prices for students subject to the Challenged Conduct to students not subject to the Challenged Conduct, while controlling for other factors that may explain Effective Institutional Price variation. It compares these groups both *before and after* the start of the Challenged Conduct, as well as *across* Defendant groups that participated and did not participate in the Challenged Conduct during the same years. For example, in the 2016 academic year, Brown, Caltech, Chicago, Emory, Johns Hopkins, and Yale had not participated in the Challenged Conduct. Therefore, these Defendants' Effective Institutional Prices serve as a benchmark for comparison of the other Defendants' Effective Institutional Prices during the same year. In contrast, when Dr. Stiroh separately runs her regressions by Defendant, she alters the underlying identification strategy of my approach, which is designed to address whether the Challenged Conduct in which all Defendants participated at some point during the Class Period had a price effect across Defendants. Instead of the regression using both the *same* Defendant and *other* Defendants as benchmarks, Dr. Stiroh's Defendant-specific regression only considers variation in Effective Institutional Prices within the same Defendant. As a result, I find Dr. Stiroh's methodology incomparable and inferior to the methodology that I proffer.

82. Dr. Stiroh also notes that there is "wide variation in the net prices paid by proposed Class members," and presents a figure showing the "Distribution of the Percentage of Cost of Attendance Covered by Gift Aid" by Class Members in the 2015 academic year.²⁰³ This is pointless. Dr. Stiroh is stating that different Class Members paid different net prices. This has no bearing on any of my analyses, and it is irrelevant to my use of a common overcharge. Class Members *must* pay

203. Stiroh Report ¶173; *id.* at Figure 7.1.

different Effective Institutional Prices—if they did not, then neither my nor Dr. Stiroh’s regressions would produce any output.

ii. My Use of a Common Overcharge Across Time Periods is Standard Practice

83. Dr. Stiroh tests splitting up my regression data and running separate regressions by time period. Her rationale is that CM guidelines had varied over time²⁰⁴ and, because of this variation, she asserts that there is no basis to assume that the Challenged Conduct would result in a uniform impact over the time frame 2003-2022.²⁰⁵

84. As an initial matter, it is unclear what Dr. Stiroh means by “uniform impact.” If she is asserting that the impact of the Challenged Conduct on each Class Member would need to be identical to be relevant, she is not addressing my conclusions. I was asked to determine whether common evidence was capable of showing some harm to each Class Member, not whether common evidence could show identical harm to each Class Member. Putting that aside, Dr. Stiroh’s argument is effectively that changes to CM guidelines over time represent uncontrolled confounders, such that, if they had appeared in my regression model, they would have affected the conduct coefficient. But Dr. Stiroh offers no analytical support that such variables truly represent omitted variables that belong in the regression. Moreover, Dr. Stiroh argues that such modifications to the CM occurred throughout the Class Period. If so, it stands to reason that some of these modifications could have occurred as a result of the effects on financial aid that the CM already exerted earlier during the Class Period. In that case, including such effects as “control” variables would be ill advised, as these would result in

204. *Id.* ¶171.

205. *Id.*

simultaneity bias—that is, when the dependent variable (Effective Institutional Price) affects an independent variable (changes to CM).²⁰⁶

85. Further, Dr. Stiroh offers no evidence that her claimed updates, revisions, and expansions of the CM guidelines affected the conduct coefficient. For her point to have any merit, she would have to show that such changes resulted in no overcharge to affected students. If changes she posits simply affected the size of the overcharge while still maintaining it, then, at best, Dr. Stiroh argues about the size of the overcharge, not its existence.

86. Instead of basing her conduct variable split dates on when changes to the CM guidelines occurred, Dr. Stiroh arbitrarily splits up my conduct variable into four separate conduct variables for subperiods 2003-2007, 2008-2014, 2015-2019, and 2020-2022.²⁰⁷ Her justification for using these specific periods is unsound. For instance, Dr. Stiroh explains that her choice to limit one of the four conduct variables to 2015-2019 is because during that period “Chicago left the 568 Group and Yale rejoined the 568 Group,” and her explanation for the 2020-2022 conduct variable is that “Vanderbilt and Penn left the Group, and Johns Hopkins and Caltech joined the Group.”²⁰⁸ Dr. Stiroh does not explain why the conduct variable would be expected to differ when one Defendant leaves the group and another Defendants enters, as her explanation describes. The econometrics literature

206. See WOOLDRIDGE at 558 (“It is useful to see, in a simple model, that an explanatory variable that is determined simultaneously with the dependent variable is generally correlated with the error term, which leads to bias and inconsistency in OLS.”).

207. Stiroh Report ¶¶175-179.

208. *Id.* n. 401.

advises against imposing arbitrary breaks such as what Dr. Stiroh proposes without having a strong *a priori* basis for doing so.²⁰⁹

b. Dr. Hill's Criticisms of My Choice of Robust Standard Errors Are Inapposite

87. Dr. Hill's criticism on my choice of robust standard errors should be understood in the limited context in which it applies. Whatever alternative choice for robust standard errors he proposes has no bearing on the economic significance of the conduct coefficient. The choices of standard errors do not change the size of the coefficient at all. As explained in my Initial Report, the conduct coefficient that I obtain in my main model (and indeed in each variant thereof) is practically (i.e., economically) significant. Dr. Hill's criticism regarding choices of standard errors does not affect the economic significance of my result.

88. Dr. Hill argues that using his preferred variant of robust standard errors (two-way clustered standard errors) robs the conduct coefficient of *statistical significance*, insinuating that lack of statistical significance indicates a zero effect. Such an inference is incorrect. Dr. Hill's argument elevates statistical significance to an undue level of importance. The American Statistical Association (ASA) Position Statement cautions:

A *p*-value, or statistical significance, does not measure the size of an effect or the importance of a result. Statistical significance is not equivalent to scientific, human, or economic significance. Smaller *p*-values do not necessarily imply the presence of larger or more important effects, and larger *p*-values do not imply a lack of importance or even lack of effect. By itself, a *p*-value does not provide a good measure of evidence regarding a model or

209. See, e.g., Bruce E. Hansen, *The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity*, 15(4) *Journal of Economic Perspectives* 117-128, 118 (2001) ("The classical test for structural change is typically attributed to Chow (1960). His famous testing procedure splits the sample into two subperiods, estimates the parameters for each subperiod, and then tests the equality of the two sets of parameters using a classic F statistic . . . However, an important limitation of the Chow test is that the breakdate must be known *a priori*. A researcher has only two choices: to pick an arbitrary candidate breakdate or to pick a breakdate based on some known feature of the data. In the first case, the Chow test may be uninformative, as the true breakdate can be missed. In the second case, the Chow test can be misleading, as the candidate breakdate is endogenous—it is correlated with the data—and the test is likely to indicate a break falsely when none in fact exists. Furthermore, since the results can be highly sensitive to these arbitrary choices, different researchers can easily reach quite distinct conclusions—hardly an example of sound scientific practice.").

hypothesis. Researchers should recognize that a p -value without context or other evidence provides limited information.²¹⁰

According to the ASA, context matters. Applied here, economic significance is equally if not more important than statistical significance.

89. Regarding Dr. Hill's alteration, clustering effectively reduces the sample size of the study. Essentially, clustering reflects the view that because the errors are related in some systematic way (e.g., within a geographic level), the sample contains less information than it would have if no such relationship had existed. That Dr. Hill observes higher standard errors with clustering reflects this sample size reduction. For this reason, clustering should be taken with care, as choosing the wrong clustering level, as Dr. Hill does, can lead to misleading inference. Moreover, the p -values for the models that purportedly "fail" in Hill's Figure 31 (models 1, 2, and 4) with clustered standard errors are 0.097, 0.077, and 0.171. Yet, as is evident from their size, two of the three p -values that he deems as "failures" are still statistically significant at a ten percent significance level. As explained in my Initial Report, I did not base my opinions on statistical evidence alone, contrary to the approaches Defendants' Experts took. Consistent with the ASA's guidance on this matter, which represents the accepted statistical practice as reflected in the literature, I leveraged not only the data but contextual factors and record evidence in supporting my opinions. The ASA recommends exactly the approach I took:

210. Ronald L. Wasserstein and Nicole A. Lazar, *The ASA's Statement on p -Values: Context, Process, and Purpose*, 70(2) THE AMERICAN STATISTICIAN 129-133, 132 (2016) [hereafter ASA Position Statement]. See also Sander Greenland et al., *Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations*, 31 EUROPEAN JOURNAL OF EPIDEMIOLOGY 337-350, 338 (2016) ("Among the many reasons are that, in most scientific settings, the arbitrary classification of results into 'significant' and 'non-significant' is unnecessary for and often damaging to valid interpretation of data; and that estimation of the size of effects and the uncertainty surrounding our estimates will be far more important for scientific inference and sound judgment than any such classification."). See also *id.* at 339 ("Much statistical teaching and practice has developed a strong (and unhealthy) focus on the idea that the main aim of a study should be to test null hypotheses.").

Researchers should bring many contextual factors into play to derive scientific inferences, including the design of a study, the quality of the measurements, the external evidence for the phenomenon under study, and the validity of assumptions that underlie the data analysis.²¹¹

90. His singular focus on statistical significance notwithstanding, Dr. Hill criticizes my specific choice of robust standard errors, the “Huber-White” errors. The relevant literature explains the shortcomings of the assumptions upon which he bases his opinion that I should have used cluster-robust standard errors instead of Huber-White standard errors. Statistical significance depends on the t-score associated with a particular variable. That t-score equals the coefficient divided by its standard error. The common standard-error calculation assumes that observations are independent, which in turn means that the error term of the regression (the difference between fitted and actual values) has a constant variance (i.e., the errors are “homoskedastic.”) Non-constant variance of the error term is known as “heteroskedasticity.” In the presence of heteroskedastic (non-constant) error terms, researchers *may* apply robust standard errors to account for existing or possible relationships between observations, the most common of which are the Huber-White standard errors. Huber-White standard errors adjust for heteroskedasticity of unknown form and enjoy widespread usage in applied and research work.²¹²

91. I emphasize that, even in the presence of heteroskedastic (non-constant) standard errors, the coefficients remain unbiased.²¹³ As noted above, the choice of whether or not to use robust standard errors or which ones to use only affects the t-score and hence the statistical significance associated with a given coefficient. The size of the coefficient, i.e., its economic (aka, practical or material) significance remains unaffected.

211. ASA Position Statement at 132.

212. WOOLDRIDGE at 271.

213. Bias is defined as “[t]he difference between the expected value of an estimator and the population value that the estimator is supposed to be estimating.” *See, e.g., id.* at 845.

92. Instead of the Huber-White standard errors, Dr. Hill contends that I should have used “clustered” (or “cluster-robust”) standard errors. Specifically, he explains:

To correct Dr. Singer’s erroneous standard errors, I replace his Huber-White robust standard errors with two-way clustered standard errors and re-estimate his model. The two-way clustered standard errors allow for correlation for a given student across years and for correlation among different students attending the same school during the same year by clustering on (1) student and (2) school by year. These clustered standard errors can measure the precision of an estimated coefficient more accurately than Dr. Singer’s standard errors.²¹⁴

The application of clustered standard errors consists of two steps. A researcher must decide (1) whether or not to cluster in the first place, i.e., whether the error structure exhibits relationship within and/or between certain groups, and (2) the appropriate level of clustering. Dr. Hill argues that clustering is warranted and that the clustering should occur at the student and by school/year levels, hence his use of two-way clustered standard errors. Dr. Hill relies, however, on two faulty arguments to support his position that my use of Huber White standard errors is unreasonable. I address each in turn.

93. Dr. Hill first mischaracterizes the nature of the Challenged Conduct, implying that each Defendant decided whether or not to join the 568 Group independently each year, akin to each institution annually tossing a coin to decide membership for the year. If Dr. Hill’s view truly reflected reality, then the probability of observing a school such as Northwestern or Duke as a member every year would be infinitesimally small, approximately equal to 0.5^t , where t equals the number of years during the relevant period in which these schools participated in the conspiracy. Dr. Hill claims:

First, Dr. Singer’s standard errors assume that all common factors that impact the Effective Institutional Price of all students at a school in the same year are explicitly controlled for in his regressions. This assumption is not credible because it assumes that there are no school-level decisions or exogenous shocks that vary over time and affect multiple students in the same school-year combination. The Challenged Conduct is one such factor that varies by school and year.

214. Hill Report ¶189.

Dr. Hill's claim runs counter to the evidence. Membership in the 568 Group represents the key feature of the Challenged Conduct and reflects the common thread that bound Defendants together both in a given year and across years. In other words, the treatment effect did not occur at the level of clustering that Dr. Hill chooses, but rather across those levels. Dr. Hill erroneously redefines the scope of the Challenged Conduct to correspond to his clustering choice. As MacKinnon et al. explain, "Thus it never makes sense to cluster at a level finer than the one at which treatment is assigned."²¹⁵

94. Abadie et al. address the very misconceptions that motivate Dr. Hill's claim that I should have used clustered standard errors rather than Huber-White.²¹⁶ The first two issues that Abadie et al. raise speak directly to Dr. Hill's misreading of the literature:

We use our framework to highlight three common misconceptions surrounding clustering adjustments. The first misconception is that the need for clustering hinges on the presence of a nonzero correlation between residuals for units belonging to the same cluster. We show that the presence of such correlation does not imply the need to use cluster adjustments. The second misconception is that there is no harm in using clustering adjustments when they are not required, with the implication that if clustering the standard errors makes a difference, one should cluster.²¹⁷

95. The guidance from Abadie et al. has direct implications for the empirical investigation of the effect of the Challenged Conduct. The authors make a critical point, which supports my decision not to employ clustered standard errors in this case:

the sampling process and the treatment assignment mechanism solely determine the correct level of clustering; the presence of cluster-level unobserved components of the outcome variable becomes irrelevant for the choice of clustering level.²¹⁸

215. James G. MacKinnon, Morten Ørregaard Nielsen, and Matthew D. Webb, *Cluster-robust inference: A guide to empirical practice*, 232(2) JOURNAL OF ECONOMETRICS 272-299, 278 (2023) [hereafter MacKinnon et al. (2023)].

216. Alberto Abadie, Guido Imbens, Susan Athey, and Jeffery Wooldridge, *When Should You Adjust Standard Errors for Clustering?*, 138(1) QUARTERLY JOURNAL OF ECONOMICS 1-35 (2023) [hereafter Abadie et al. (2023)].

217. Abadie et al. (2023) at 3.

218. *Id.* See also *id.* at 32 ("If assignment is perfectly clustered so that units that belong to the same cluster have the same treatment assignment, there is no improvement from using the CCV [cluster causal variance] variance and the TSCB [two-stage cluster bootstrap] variance estimator is not applicable.").

In other words, the sole determinant of whether one should cluster or not depends on the level of the treatment assignment. In this case, the treatment assignment (exposure to the Challenged Conduct) did not vary by school or year. Dr. Hill’s approach seeks to manufacture variation where none exists. The point of the Challenged Conduct was to impose a common financial aid approach, and to reduce variability among schools.²¹⁹ Contrary to Defendants’ own admissions to this effect, Dr. Hill posits the exact opposite—namely, that membership in the 568 Group varied across schools and years. Dr. Hill cites to MacKinnon et al., who explain that “When the regressor of interest is a treatment dummy, and the level at which treatment is assigned is known, then it generally makes sense to cluster at that level.” I agree. But because the treatment assignment occurred across schools and years, it does not make sense to adopt the by-school-and-year clustering that Dr. Hill proposes.

96. In addition, the decision of whether or not to cluster, as Abadie et al. explain, depends on “the sampling process”—that is, how we obtained the data, what those data represent, and what role the data play in a researcher’s analysis. I understand that these data represent the full available complement of data for the population of students for each Defendant in the relevant period. I do not seek to extrapolate from these Defendants to other schools not found in the data.

97. In a discussion of Abadie et al.’s paper on the World Bank blog, lead economist David McKenzie explains that “Instead, under the sampling perspective, what matters for clustering is how the sample was selected and whether there are clusters in the population of interest that are not represented in the sample.”²²⁰ Abadie et al. further note that, “when the number of clusters in the sample is a nonnegligible fraction of the number of clusters in the population, conventional clustered

219. See, e.g., *Memorandum of Understanding*.

220. David McKenzie, *When should you cluster standard errors? New wisdom from the econometrics oracle*, WORLD BANK BLOG (Oct. 16, 2017), <https://blogs.worldbank.org/en/impactevaluations/when-should-you-cluster-standard-errors-new-wisdom-econometrics-oracle> (“Their advice [Abadie et al. (2023)]: whether or not clustering makes a difference to the standard errors should not be the basis for deciding whether or not to cluster. They note there is a misconception that if clustering matters, one should cluster.”).

standard errors can be severely inflated...”²²¹ In this case, the number of clusters in the sample data equal the number of clusters in the population.

98. Dr. Hill inaccurately claims that I “assume that all common factors that impact the Effective Institutional Price of all students at a school in the same year are explicitly controlled for in his regressions.”²²² I make no such assumption, and I reject Dr. Hill’s apparent implication of omitted variable bias. In addition to the treatment variable of interest, the Challenged Conduct, I controlled for any potential factors that could confound the relationship between the treatment and outcome variables. As such, my identification strategy aimed at recovering a causal effect from the treatment on the outcome. I did not seek to maximize the fit of my regression, i.e., to explicitly include all variables that could affect the institutional price students paid at a given school in a given year. Nevertheless, my regression explained the vast majority of the variation in such prices. Moreover, such as assumption is unnecessary. As Abadie et al. noted above, “the presence of cluster-level unobserved components of the outcome variable becomes irrelevant for the choice of clustering level.”²²³

99. In his second critique, Dr. Hill argues:

Dr. Singer’s standard errors assume that there are no factors that are not explicitly controlled for in his regressions that impact the Effective Institutional Price of a student at a school across years. This assumption is not credible because, for it to be true, it would have to be the case that any information at the student level that is not explicitly controlled for in Dr. Singer’s model (e.g., whether the student’s parents are married or divorced, any home equity owned by the student’s parents, etc.) is completely random from year to year. This is clearly not the case. Such correlation over time is called serial correlation, and it is standard to correct for it when calculating standard errors.²²⁴

Dr. Hill misrepresents my analysis. I specifically control for factors he identifies above. First, I include student fixed effects, which capture any time-invariant student-level characteristics *across*

221. Abadie et al. (2023) at 1-2.

222. Hill Report ¶186.

223. Abadie et al. (2023) at 3.

224. Hill Report ¶187.

the years. Even if some between school-year correlation existed for a student, including student fixed effects would capture at least a portion, though likely not all of this correlation.²²⁵ Second, I include student-specific characteristics that may vary over time, such as parental net worth as well as other key financial metrics. In addition, I control for the student's year in school. As such, while a student's characteristics may be similar across years, I reject Dr. Hill's assumption that the prediction errors must be correlated in this fashion. To wit, Dr. Hill provides no support for his claim of serially correlated errors. On the contrary, Dr. Hill's only justification appears to be that his clustered standard errors are higher than the Huber-White standard errors that I use. This logic reflects the exact misconception against which Abadie et al. warn.

c. Contrary to Dr. Hill's Assertions, My Overcharge Models Are Robust to His Data-Processing and Control-Variable Adjustments

100. In Section 8.2 of his report, Dr. Hill claims that I erred in both my processing of Defendants' data,²²⁶ as well as in my construction of control variables.²²⁷ For ease of exposition, I will use the terminology "data processing" to refer to the process that my team had applied to cleaning Defendants' structured data. I will use the terminology "control variable construction" to refer to any data adjustments that my team had applied to Defendants' data after aggregating it into a stacked, regression dataset.²²⁸ Dr. Hill makes a substantial number of adjustments to both my data and control variable construction methodology, and he asserts that these adjustments produce more accurate results, which he claims show no reliable evidence of overcharges.²²⁹

225. MacKinnon et al. (2023) at 277.

226. Hill Report §8.2.1.

227. *Id.* §8.2.2.

228. Control variable construction also refers to the inclusion of IPEDS data into my regression analysis since IPEDS control variables were added to the dataset after Defendants' data had been aggregated.

229. Hill Report §8.2.4.

101. I explain below that the vast majority of Dr. Hill's data-processing and control-variable construction critiques either have no practical effect on my results or are unwarranted. I accept certain adjustments that Dr. Hill makes, and I dismiss others that I find to be improper. After making the adjustments that I concede are proper, my regressions continue to show a positive, economically, and statistically significant conduct coefficient, contrary to Dr. Hill's claims. Therefore, the adjustments that Dr. Hill expounds in Section 8.2 of his report do not alter my opinion that the Challenged Conduct resulted in an artificial inflation in Effective Institutional Prices, nor do these adjustments alter my opinion that the artificial inflation in Effective Institutional Prices affected all or nearly all Class Members.

102. Dr. Hill makes a *substantial* number of adjustments to both my regression data and my methodology, yet he never shows the *individual* effect of making each of these separate adjustments. Rather, Dr. Hill presents the *cumulative* impact of including *every* one of his adjustments on my regression results in his Figure 33. His Figure 33 shows how my regression changes after: (1) adjusting my data for every data-processing critique specified in his report, as well as adjusting for numerous other processing critiques that he outlines in a Microsoft Excel document provided in his backup; (2) adjusting my data to correct for what he claims are purported flaws in how I construct my control variables; and (3) adjusting my standard errors to be two-way clustered standard errors (discussed above in Part III.A.2.b.). It is not surprising that Dr. Hill finds that his myriad adjustments alter my regression results—he cumulatively makes over 60 adjustments in total to both my data and methodology.²³⁰ Yet, by applying all adjustments at once, Dr. Hill's analysis obfuscates which adjustments are relevant to his downgrading of my findings.

230. See Table 5 and Table 5 note, *infra*.

103. In Table 5, I outline every adjustment that Dr. Hill makes to my data-processing and control-variable construction in Section 8.2 of his report. I provide an explanation of the significance of making each of these adjustments in isolation, and I explain whether I accept the adjustment. Most of the adjustments that Dr. Hill proffers and that I accept have no effect or a very modest effect on my regression data and on my results. The column “Summary of Effect” is meant to provide a general view of whether the adjustment has any meaningful impact on the regression data based on my own personal judgment, and the column “Description of Adjustment Effect/Relevance” provides more detail as to the how I measure the effect of each change. I find that 52 of the 60 total data adjustments proffered by Dr. Hill individually and that I accept have no appreciable impact on my findings—that is, the effects are deemed trivial or modest as opposed to appreciable. In the sections below, I provide more explanation as to the relevance of Dr. Hill’s adjustments to my data-processing and control-variable construction, and I highlight which of his adjustments are the primary drivers of the discrepancy in his results and my own.

TABLE 5: DR. HILL'S MYRIAD DATA-PROCESSING AND CONTROL-VARIABLE CONSTRUCTION
ADJUSTMENTS

Overall Issue #	Defendant Affected	Adjustment Description	Accept?	Description of Adjustment Effect/Relevance	Summary of Effect
1	Brown	Dr. Hill converts observations where aid_year = "NULL" to have aid_year equal to other observations in the same dataset.	Y	Approximately 99 percent of observations where aid_year = "NULL" have missing income or asset data. Therefore, approximately 99 percent of these observations end up being excluded from the regression anyways.	Trivial
2	Brown	Dr. Hill adjusts my code to no longer drop both duplicates of same student in same year.	Y	This change only affects 15 students.	Trivial
3	Brown	Dr. Hill adjusts how I code Brown's academic years. He treats aid_year as referring to the spring semester instead of the fall semester.	Y	This change affects all observations since it converts academic year to academic year minus one.	Appreciable
4	Caltech	Dr. Hill incorporates additional awards data for 2020-2022.	Y	There are only 653 additional observations in the awards data that Hill uses versus the awards data that I used.	Trivial
5	Caltech	Dr. Hill adjusts my code to include student net worth in the family net worth total.	Y	Parental net worth is the primary component of family net worth (student net worth comprises less than 2 percent of family net worth for Caltech).	Trivial
6	Chicago	Dr. Hill recategorizes all institutional gift aid as being need-based, other than 9 specific merit scholarships listed online.	Y	This change increases the conduct coefficient of my primary Effective Institutional Price regression by 45.	Modest
7	Chicago	Dr. Hill adjusts my code to include the IM parents' business value variable in some instances where I did not include this variable.	Y	This change only affects one component of the parental net worth variable, and it affects this variable for less than 7 percent of observations.	Trivial
8	Chicago	Dr. Hill incorporates 2016-2022 financial aid data for Chicago.	N	The change in Chicago's database system renders these data incompatible with the earlier data. <i>See</i> Part III.A.2.c.	N/A
9	Columbia	Dr. Hill incorporates additional awards data which covers COVID grants.	Y	This change increases the "other grants awarded" from \$59.9 million to \$108.1 million. However, the other grants awarded are only 5.5 percent of Columbia's Financial Aid awards and thus this change is modest	Modest

Overall Issue #	Defendant Affected	Adjustment Description	Accept?	Description of Adjustment Effect/Relevance	Summary of Effect
10	Columbia	Dr. Hill adjusts my award cleaning code to take the sum of awards per student-year, rather than preserving the highest value award per student-year.	Y	While this change materially increases the amount of institutional grant aid awarded between 1998 and 2010, it does not materially alter the variation in Columbia's need-based institutional grant aid—I find that Dr. Hill's real need-based institutional grant aid variable has a correlation with my version of this variable of 99.8 percent for those students that show up in both datasets.	Appreciable
11	Cornell	Dr. Hill replaces two files that I had relied upon with new versions that I did not have access to at the time of my Initial Report submission.	Y	This change pertains to admissions data, and it therefore has no bearing on my findings.	None
12	Cornell	Dr. Hill incorporates the Cornell statutory COA for New York State residents.	Y	This change increases the conduct coefficient of my primary Effective Institutional Price regression by 98 dollars.	Modest
13	Dartmouth	Dr. Hill subtracts 1 from the academic year if the term_code field suggests it is a spring term.	Y	This changes the year for approximately 55 percent of observations. This does not substantively affect my results.	Trivial
14	Dartmouth	Dr. Hill drops award designated "DO NOT USE" in the award description.	Y	This change only impacts approximately 0.2 percent of observations. Dr. Hill also never confirms whether this designation implies the student did not receive an award.	Trivial
15	Emory	Dr. Hill adjusts my code code to not exclude all duplicated observations of the same student in the same year.	Y	This change only pertains to the admissions data, and it therefore has no bearing on my findings.	None
16	Emory	Dr. Hill recategorizes BBA and Education Abroad to be need-based aid, rather than merit-based aid.	N	Dr. Hill incorrectly recategorizes these awards. He initially converts them to be need-based, but then later in his code reconverts 80 percent of them back to being merit-based. These awards comprise a trivial proportion of all awards.	N/A
17	Emory	Dr. Hill incorporates a supplemental production for purposes of categorizing financial aid award types.	Y	This change results in a difference between Dr. Hill's version of total institutional need based grant aid and my version of total institutional need based grant aid of less than 0.1 percent.	Trivial
18	Duke	Dr. Hill adjusts the academic year based on whether the field "admittermdescription" specifies spring, summer, or fall.	Y	This change pertains to admissions data, and it therefore has no bearing on my findings.	None

Overall Issue #	Defendant Affected	Adjustment Description	Accept?	Description of Adjustment Effect/Relevance	Summary of Effect
19	Duke	Dr. Hill adjusts my award cleaning code to include awards with a typo ("Scholarhp") to be treated as merit-based aid.	Y	While this change impacts 44 percent of Duke students during the affected years, it only affects their merit aid, and not their need-based aid. It therefore does not affect the dependent variable of my primary regression. Rather, it indirectly affects my model results via its impact on the "other grant aid" control variable and by its effect on the sample selection.	Appreciable
20	Duke	Dr. Hill adjusts my award cleaning code to include "scholarships" in federal, state, and other gift aid.	Y	This change increases federal grant awards by approximately \$1.3M (or approximately 0.8 percent of the total value).	Trivial
21	Duke	Dr. Hill adjusts the academic year by subtracting one to indicate that the aid year field represents the spring of an academic year.	Y	Running my primary regression model using Duke's data adjusted to one year prior (as Hill does) results in the conduct coefficient decreasing from 1,497 to 1,180.	Appreciable
22	Duke	Dr. Hill removes the awdperiod "N", which represents summer terms.	Y	The summer awdperiod makes up approximately 3.1 percent of the institutional grant aid awarded and 0.9 percent of the federal grant aid awarded.	Modest
23	Duke	Dr. Hill replaces my use of "accepted" amounts with "paid" amounts.	N	The "paid" amount variables are not accurate. For instance, named plaintiff Sia Henry's award amounts do not equal the sum of her "paid amounts" but do equal the sum of her "accepted" amounts.	Appreciable
24	Georgetown	Dr. Hill incorporates additional data covering Federal and outside awards.	Y	These additional data comprise less than 7 percent of all awards (of all types). It bears noting though that this change results in an additional \$80 million of institutional grant aid awards being accounted for in 2023 and an additional \$500 thousand awarded in 2024. This should produce a reduction in effective institutional prices relative to my original data. Because 2023-2024 are the two "clean" years for Georgetown, this results in the model attributing greater institutional grant aid to the "clean" period.	Modest
25	Georgetown	Dr. Hill recategorizes certain awards, such as by dropping graduate awards.	Y	All of Hill's adjustments to Georgetown's data result in less than a four percent difference in total financial aid.	Modest
26	JHU	Dr. Hill recategorizes some awards that I had categorized as "Institutional" to "Federal."	Y	This change affects less than 0.2 percent of awards observations.	Trivial

-71-

Overall Issue #	Defendant Affected	Adjustment Description	Accept?	Description of Adjustment Effect/Relevance	Summary of Effect
27	JHU	Dr. Hill recategorizes some awards that I had classified as need-based to instead be merit-based.	Y	This change affects approximately 0.5 percent of observations. It bears noting that Dr. Hill's change is inconsistent with JHU's "meritaward" variable, which shows some of the awards that Dr. Hill classifies as being "merit-based" as having a "meritaward" value of "N".	Trivial
28	JHU	Dr. Hill drops graduate awards from the awards data.	Y	This change affects 46 observations (0.02 percent of all observations).	Trivial
29	JHU	Dr. Hill claims that my code results in duplicate observations across the files for the 2023-2024 academic year. He adjusts my code to exclude duplicates.	Y	This change affects 6.1 percent of the observations in the awards data.	Modest
30	MIT	Dr. Hill adds yellow ribbon and veterans benefits from the long award data.	Y	This change affects only ~0.01 percent of observations. It bears noting that Dr. Hill does not provide documentation to support his assumptions that the fund codes "VATUTN", "YELRBN", "VASTIP", and "YELMCH" correspond to what he claims they are.	Trivial
31	Notre Dame	Dr. Hill incorporates an award type crosswalk into my analysis.	Y	This change increases institutional grant aid in the final dataset by 66.1 percent. Prior to the production of these data, I had to categorize institutionally sourced aid from non-institutionally sourced aid based on the fund names, which required assumptions based on terminology used in the fund name (e.g., any fund containing "ND" is institutional).	Appreciable
32	Notre Dame	Dr. Hill adjusts my code to use "aidy_code" instead of "activity_date" to identify the academic year.	Y	Of the 126,753 uid/aidyear pairs in the Singer Notre Dame FA Dataset, only 88.6 percent of the uid/aidyear pairs have a matching pair in Dr. Hill's Notre Dame FA Dataset. This implies that around 11 percent of observations were affected by this change	Modest
33	Northwestern	Dr. Hill replaces some of the datasets that I had used with newer versions produced later on.	Y	This change affects approximately 0.3 to 3 percent of observations within each file that was updated.	Trivial
34	Northwestern	Dr. Hill adjusts the academic year by subtracting one to indicate that the aid year field represents the spring of an academic year.	Y	Running my primary regression model using Northwestern's data adjusted to one year prior (as Hill does) results in the conduct coefficient decreasing from 1,497 to 1,459.	Modest

PRIVILEGED AND CONFIDENTIAL
PREPARED FOR COUNSEL

Overall Issue #	Defendant Affected	Adjustment Description	Accept?	Description of Adjustment Effect/Relevance	Summary of Effect
35	Northwestern	Dr. Hill adjusts my code to include awards with type = "A" and category = "NU Athletic scholarships" as being institutional merit-based aid.	Y	This change affects approximately 0.8 percent of observations. My categorization of these awards was based on their "source" being listed as "O" (which could correspond to "outside" or "other").	Trivial
36	Northwestern	Dr. Hill categorizes all awards with source = "I" and missing type as need-based, institutional grants.	Y	This change affects approximately 0.02 percent of observations.	Trivial
37	Northwestern	Dr. Hill recategorizes some awards that are classified as Pell Grants or Fed ACG awards as being Federal/state awards.	Y	This change affects approximately 0.24 percent of observations. It bears noting that these awards were recorded in the data as having "source" equal to "I" (for institutional). Dr. Hill's adjustment therefore contradicts the data "source" field for these awards.	Trivial
38	Northwestern	Dr. Hill removes 529 plans, family support, prepaid tuition plans, and unmet need from award totals.	Y	This change affects approximately 0.48 percent of observations.	Trivial
39	Penn	Dr. Hill incorporates an award crosswalk (PENN568_STRUCTURED-00000216) that was produced in his workpapers, which I did not have access to. I therefore could not definitively identify institutional grant awards.	Y	These changes increase the amount of institutional grant awards by 60.4 percent. However, this change increases Penn's institutional grant awards by a weighted average of 68.5 percent during the years Penn was outside of the cartel (1998-2002 and 2020-2021) and increases the institutional grant awards by a weighted average of 58.4 percent for the years within the cartel (2003-2019).	Appreciable
40	Penn	Dr. Hill adjusts my code to include other, outside grants in Penn's data prior to 2022.	Y	The other, outside grant awards that Dr. Hill highlights only account for 1.3 percent of all awards in the pre-2022 financial aid data.	Trivial
41	Penn	Dr. Hill recategorizes Mayors Scholarships and Named Scholarships from merit-based to need-based awards.	Y	Incorporating Dr. Hill's adjustment to recategorize these awards causes Penn's institutional grant aid to increase by 55 percent in 2022 and by 136 percent in 2023. This would be expected to increase the conduct coefficient, since it attributes greater institutional grant aid to years outside of the conduct period.	Appreciable
42	Penn	Dr. Hill adds veteran benefits to Federal grant aid.	Y	Out of the 12,201 observations in the combined 2022 and 2023 FA datasets, only 7 observations (0.06 percent of total) were paid veterans benefits.	Trivial

-73-

Overall Issue #	Defendant Affected	Adjustment Description	Accept?	Description of Adjustment Effect/Relevance	Summary of Effect
43	Penn	Dr. Hill deduplicates the data after incorporating the most recent CIRQUE analytics UID crosswalk.	Y	This data duplication issue has almost no effect. For example, Hill's deduplication in the penn568-structured-00000115 file only drops 2 of the 111,723 observations. For penn568-structured-00000116, this deduplication drops 11 of the 111,741 observations	Trivial
44	Rice	Dr. Hill adjusts how I go about filling in Rice's missing 2021 awards data. I had relied on another file that Rice produced as a supplement to fill in Rice's missing 2021 awards. Dr. Hill claims that the supplemental data I used are incomplete, and he instead imputes the missing 2021 values using a different dataset.	Y	The 2021 data I had used only contained first years. Dr. Hill's adjustment therefore increases the 2021 aggregate awards by approximately four-fold.	Modest
45	Rice	Dr. Hill adjusts my code to include athletic scholarships in the institutional grant aid total.	Y	Athletic scholarships account for 5 percent to 23 percent of institutional grant awards from 2009-2023.	Modest
46	Rice	Dr. Hill adjusts my code to use different variables for student and parental net worth, rather than the variable that I had used, which he claims only include net worth of investments.	Y	The student net worth variables that I use are identical to the versions that Dr. Hill uses. The parental net worth variables that I use have a correlation of 99.9 percent with the parental net worth variables that Dr. Hill uses.	Trivial
47	Vanderbilt	Dr. Hill adjusts my code to correct for what he claims are incomplete Federal, state, and outside grant categorizations.	Y	These changes (as well as the double counting fix in Issue 3) decreases the amount of federal grant aid by 13.9 percent, decreases the amount of total aid by 11.7 percent, and increases the "other grants paid" from \$0 to \$54.8 million.	Modest
48	Vanderbilt	Dr. Hill removes 529 plans and "Vandy plan" payments from the awards data.	Y	This has no effect on my results. 529 plans and "Vandy plan" payments are not included in the "disbursed amounts" column and are not included as "awarded" in my regression data.	None

-74-

Overall Issue #	Defendant Affected	Adjustment Description	Accept?	Description of Adjustment Effect/Relevance	Summary of Effect
49	Vanderbilt	Dr. Hill claims that I double count some institutional awards as being both need-based and merit-based. Dr. Hill adjusts my code by using the "need_based" field to make this distinction.	Y	Both changes 49 and 50 to non-need based aid increased the level of non-need based aid from around \$2 million to over \$463 million.	Modest
50	Vanderbilt	Dr. Hill categorizes athletic scholarships and tuition benefit awards for Vanderbilt as institutional merit-based aid.	Y	Both changes 49 and 50 to non-need based aid increased the level of non-need based aid from around \$2 million to over \$463 million.	Modest
51	Vanderbilt	Dr. Hill adjusts the academic year by subtracting one to indicate that the aid year field represents the spring of an academic year.	Y	Running my primary regression model using Vanderbilt's data adjusted to one year prior (as Hill does) results in the conduct coefficient increasing from 1,497 to 1,544.	Modest
52	Yale	Dr. Hill recategorizes awards with "Counselor" in the title or "George H. Bunker Schp" or "Yale Yellow Ribbon Schp" as non-need-based.	Y	These three award types only comprise 0.3 percent of all award observations. It bears noting that Yale's website expressly states that Yale does not award merit-based scholarships (https://finaid.yale.edu/award-letter/financial-aid-terminology/merit-based-scholarships).	Trivial
53	Yale	Dr. Hill drops awards written as "Horstmann Scholarship" (graduate award) and with "DO NOT USE" in their title.	Y	These awards comprise only 4 out of over 350K award observations.	Trivial
54	Yale	Dr. Hill claims that I exclude offer_amt_SCHLCIPE, offer_amt_SCHLCURR and offer_amt_SCHLSPEC from my total institutional awards field but that I include them in my institutional grants field. He adjusts my code to include them in my total institutional awards field.	Y	This only affects the total institutional award variable, which has no impact on any of my analyses.	None
55	All	Dr. Hill winsorizes adjusted gross income, thereby recoding the top and bottom 1 percentile values across the entire dataset.	N	See Part III.A.2.c.ii.	N/A
56	All	Dr. Hill winsorizes net worth, thereby recoding the top and bottom 1 percentile values across the entire dataset.	N	See Part III.A.2.c.ii.	N/A

PRIVILEGED AND CONFIDENTIAL
PREPARED FOR COUNSEL

Overall Issue #	Defendant Affected	Adjustment Description	Accept?	Description of Adjustment Effect/Relevance	Summary of Effect
57	All	Dr. Hill adjusts my code to incorporate IPEDS 1997 institutional tuition revenues (which enter my regressions in the lagged 1998 institutional tuition revenues control variable).	Y	This has no material effect on my results. <i>See</i> Part III.A.2.c.ii.	Trivial
58	All	Dr. Hill adjusts my coding of the year in college variable, which he claims is inconsistently coded.	Y	Using Hill's year in college increases the conduct coefficient by 4 dollars.	Trivial
59	All	Dr. Hill splits my continuous versions of the control variables for year in college and number in college into separate dummy variables.	N	<i>See</i> Part III.A.2.c.ii.	N/A
60	All	Dr. Hill alters how negative and zero values enter my logarithmic regressions by coding these variables as zero and entering a dummy variable for when they are missing.	N	<i>See</i> Part III.A.2.c.ii.	N/A

i. Dr. Hill's "Corrections" to My Data Processing Do Not Materially Alter My Findings

104. Dr. Hill critiques my processing of Defendants' structured data. He outlines numerous purported flaws in my data processing in Section 8.2.1 of his report, plus many more in his backup production. Below, I provide details as to the cause and the significance of a number of these purported flaws. After adjusting my data to incorporate the points that Dr. Hill raises and with which I agree, and to exclude the points that Dr. Hill raises with which I disagree, I find that his adjustments do not materially alter my findings, and that my regression models continue to show that the Challenged Conduct resulted in an economically and statistically significant artificial increase in Effective Institutional Prices.

105. Many of the “data processing” critiques raised by Dr. Hill stem from the inherent challenges associated with interpreting Defendants’ data, particularly given its volume and complexity. Many Defendants did not provide sufficient information to interpret many of their data fields, and so my team had to make certain assumptions regarding how to process Defendants’ data. These assumptions were reasonable given the limited nature of the disclosures relating to the data, and any adjustments that Dr. Hill made due to these assumptions has little material impact, as shown in Table 5 above.

106. Moreover, certain data-processing issues raised by Dr. Hill contradict Defendants’ data or responses provided by Defendants. For instance, Dr. Hill claims that Brown’s aid year variable refers to the spring semester, rather than the fall semester, and that Brown’s data should therefore be coded based on their aid year minus one.²³¹ This means that Brown’s data starts in the 1997 academic year, one year prior to the agreed-upon beginning date for the production of Defendants’ structured data. Additionally, over the course of correspondence that my team had with Brown (through Counsel), Brown had never specified that the aid year variable should correspond to the spring semester, even though my team had asked questions pertaining to this variable.²³² Similarly, Dr. Hill claims that I improperly dropped Brown observations that had aid year values equal to “NULL,” and that these observations should have instead been coded based on other observations within their datasets.²³³ My team had asked this exact question as to how to treat these “NULL” observations in January 2024. Brown had responded: “We do not believe that any available data is missing,” and

231. Hill Report ¶193; Point 1; Bullet 2. Over the course of question-and-answer correspondence between my team and Brown (through Counsel), Brown had never specified that the “aid_year” variable corresponds to the spring of the academic year.

232. For example, *see* Brown’s Responses to Questions re: Structured Data (Jan. 26, 2024), 7 (Q: “Is there a reason why there are only ~3,000-4,000 observations per year from 1998 – 2006, and ~14,000 from 2007?” To this, Brown responded “...Although we cannot pinpoint a precise explanation...the available 1998-2006 data...”. Brown specifically mentioned “1998-2006” when referencing the data we asked about, and did not correct it to say “1997”).

233. Hill Report ¶193; Point 2; Bullet 3.

stated they would review the problem further.²³⁴ Although my team never received an answer to this question, at a July 31, 2024 meeting with Bates White and NERA, Brown instructed Bates White as to how to interpret these “NULL” values.²³⁵

107. Dr. Hill claims that I improperly classified some of Northwestern’s federal awards as being institutional.²³⁶ Yet Northwestern’s data had specified these awards as being institutionally sourced, in contrast to Dr. Hill’s claim.²³⁷ Similarly, Dr. Hill claims that I improperly categorized some Northwestern awards that are institutional athletic scholarships as being outside scholarships.²³⁸ Yet the data to which Dr. Hill is referring contains observations where the award type is equal to “A” (athletic) and the source type is equal to “O” (other), contrary to his claim.²³⁹ Dr. Hill asserts that I incorrectly classified some Yale awards that are merit-based as being need-based.²⁴⁰ This contradicts Yale’s website, which states that “Yale does not award merit-based scholarships.”²⁴¹

234. See Brown’s Responses to Questions re: Structured Data (Jan. 26, 2024), at 5. (Q: “Why are ~25% of observations for the “aid_year” field equal to “NULL”? Can you provide the missing years?” To this, Brown responded: “We do not believe that any available data is missing. Brown has not been able to determine the reason for the missing “AID_YEAR” data.”).

235. Hill Workpapers, “Summary of data processing updates.xlsx”.

236. Hill Report ¶193; Point 3; Bullet 2.

237. Dr. Hill claims that I miscategorized some Pell Grant and “Fed ACG” awards in Northwestern’s data as being from an institutional private source. See Hill Workpapers, “Summary of data processing updates.xlsx”. It bears noting that Northwestern’s data was incorrect in its classification of these awards that Dr. Hill identifies. I use the variable “source” in Northwestern’s data to classify award sources, and I classified awards with “source” = “I” as institutional. The awards that Dr. Hill is highlighting have a “source” = “I”, inconsistent with Dr. Hill’s description. Nevertheless, I accept Dr. Hill’s adjustment of my classification for these observations, and I find that it has a very modest impact on my results (this adjustment affects less than 0.3 percent of all Northwestern observations).

238. See Hill Workpapers, “Summary of data processing updates.xlsx”.

239. The data that Dr. Hill is referring to contains observations where the award type is equal to “A” (athletic) and the source type is equal to “O” (other), yet the category is “NU Athletic scholarships”. There is a clear discrepancy between the source type column and the category column. It is therefore not clear whether Dr. Hill’s adjustment to classify these as institutional athletic scholarships is correct, or whether my classification of these awards as other grants is correct. Because these observations have no meaningful impact on the results (Dr. Hill’s adjustment only affects 0.8 percent of observations), I accept Dr. Hill’s adjustment under the assumption that the category variable is more accurate for these observations, although there is no strong basis to this assumption.

240. See Hill Workpapers, “Summary of data processing updates.xlsx”.

241. *Financial Aid: Merit-Based Scholarship*, YALE, <https://finaid.yale.edu/award-letter/financial-aid-terminology/merit-based-scholarships> (last visited Oct. 2024).

108. Some of the data-processing issues that Dr. Hill raises arise from the fact that Defendants had not yet produced the data necessary for my analyses, but later produced the data. For example, Dr. Hill uses a Penn crosswalk for classifying awards, but I did not have access to said crosswalk.²⁴² Nonetheless, in my revised estimates, I take account of these data that had previously not been available.

109. Other data-processing issues that Dr. Hill raises have no practical impact on either the data or my findings, but nonetheless I take account of them in my revised estimates. For instance, Dr. Hill claims that, for Emory and Brown, I drop all instances of a student in a year where the student shows up in multiple observations for the same year (i.e., duplicate observations). Dr. Hill claims this is improper because I should have at least preserved one instance of the duplicate student in the year, rather than dropping *all* instances.²⁴³ As Dr. Hill states in his report, this adjustment drops a trivial 15 students for Brown. And, contrary to what Dr. Hill states in his report, this adjustment has *no* effect for Emory because it pertains only to admissions data, which does not have any impact on any of my analyses. As further examples, Dr. Hill claims that I did not properly include state awards in my “other gift aid” control variable for Georgetown.²⁴⁴ State grants account for zero percent of grants at Georgetown and therefore have no material impact on my results.²⁴⁵ Dr. Hill claims that I incorrectly omitted yellow-ribbon and veteran-benefit awards in my processing of MIT data.²⁴⁶ His

242. See Hill Workpapers, “Summary of data processing updates.xlsx” (referencing PENN568_STRUCTURED-00000216).

243. Hill Report ¶193; Point 2; Bullet 4.

244. *Id.* ¶193; Point 3; Bullet 3.

245. See my workpapers for details.

246. Hill Report ¶193; Point 2. Dr. Hill claims that I incorrectly omitted yellow ribbon and veteran benefit awards in my processing of MIT data. Hill Workpapers, “Summary of data processing updates.xlsx”. Dr. Hill appears to assume that the values “VATUTN”, “YELRBN”, “VASTIP”, and “YELMCH” for the variable “fund_name” represent these purported award types. He offers no explanation or documentation to support his assumption that these values of “fund_name” should be included.

adjustment has no impact on my results—the yellow-ribbon and veteran-benefit awards that he flags are equal to zero for approximately 99.9 percent of all observations.

110. One of the important changes that Dr. Hill makes to my data processing with which I disagree pertains to the University of Chicago’s data. Chicago produced financial aid data from two database systems. Chicago stored its 1998-2015 financial aid data in a PowerFAIDS database and stored its 2016-2023 financial aid data in a PeopleSoft database.²⁴⁷ Because of this change in database systems over time, my team had reviewed and analyzed the consistency of these two databases to determine whether both databases recorded the variables used for my analyses consistently across time. Upon this review, my team found that Chicago did not maintain consistency in how it recorded data in PeopleSoft relative to how it recorded these data in PowerFAIDS. Specifically, Chicago’s 2016-2023 PeopleSoft data does not allow one to isolate institutional need-based grant aid due to issues in identifying the source of the financial aid (e.g., institutional aid, Federal aid, etc.) and in

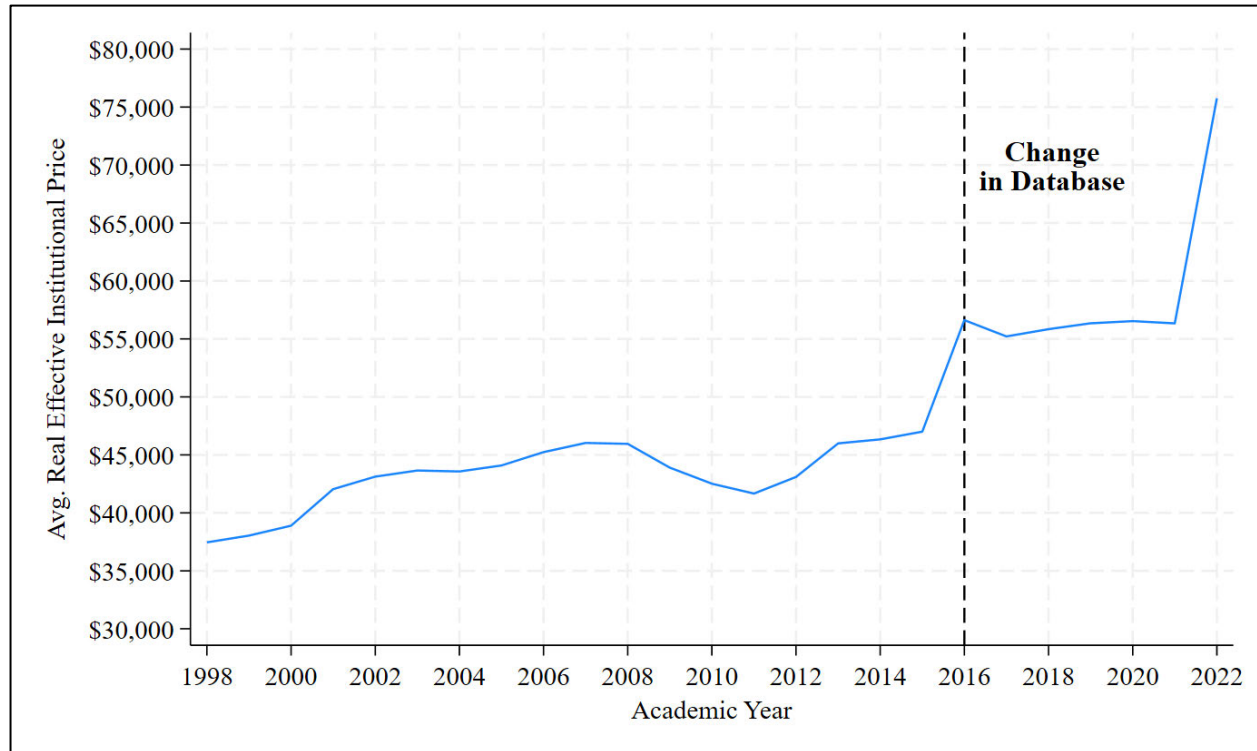
247. PowerFAIDS is “a comprehensive software solution automating financial aid processes” produced by the College Board. *PowerFAIDS*, COLLEGE BOARD, <https://powerfaids.collegeboard.org/media/pdf/powerfaids-financial-aid-management.pdf> (last visited Sept. 2024). The software offers a range of functionalities, including the ability to produced need analyses, manage federal funds, and award and simulate student aid awards. *See What PowerFAIDS Helps You Do*, COLLEGE BOARD, <https://powerfaids.collegeboard.org/about-powerfaids/what-powerfaids-helps-you-do> (last visited Sept. 2024). PeopleSoft Campus Solutions is a similar software produced by Oracle. It “automates federal and institutional financial aid processing for a more efficient operation. It provides flexibility and helps you manage financial aid activity for applicants and students.” *Campus Solutions 9.2: Financial Aid*, ORACLE, https://docs.oracle.com/cd/F33383_01/psft/pdf/cs92lsfa-b072020.pdf (last visited Sept. 2024), 39. The PeopleSoft database allows an institution to: “Receive and track financial aid applications. Design cost of attendance assessment by defining budget categories, items, and formulas. Manage need analysis, packaging, disbursement, and loan processing. Process origination and disbursement for FFELP, Direct Loan, and Pell Grant programs. Calculate federal Pell awards. Use award plans to automatically package students. Match financial aid sources to eligible students. Modify financial aid notification letters. Maintain federal compliance with annual regulatory updates for ISIR, Pell, Direct Loan, FISAP, and other federal updates. Post financial aid to student accounts. Provide students with self-service access to view, accept, and decline awards. Manage student work-study programs.” *Id.* at 39.

distinguishing need-based from merit aid types.²⁴⁸

111. In Figure 1, I demonstrate this inconsistency in Chicago's databases by plotting Chicago's average real Effective Institutional Price from 1998 to 2022. These data come directly from Dr. Hill's backup. The vertical dashed line shows when Chicago had switched from a PowerFAIDS to a PeopleSoft database system. As shown, there is a significant increase in Chicago's average real Effective Institutional Price from approximately \$47,000 per academic year to \$57,000 per academic year when Chicago switched its database system—amounting to a \$10,000 increase year-over-year. A change in real Effective Institutional Prices such as that shown in Figure 1 is attributed to the change in database systems. Figure 1 also shows a substantial increase in real Effective Institutional Prices from approximately \$56,000 to over \$76,000 between the 2021 and 2022 academic years—an \$19,000 increase in real Effective Institutional Prices, which are *after* accounting for inflation. This spike in real Effective Institutional Prices in a single year is implausible and likely reflects another inconsistency in Chicago's database.

248. Dr. Hill highlights that I did not use Chicago's 2016-2023 data. Hill Report ¶193 (at point 2, bullet 1).. I did not use Chicago's 2016-2023 data due to problems identifying institutional grant aid and problems with identifying other, non-institutional grant aid. Chicago had produced two datasets containing awards for the period 2016-2023: "UCHICAGO_STRUCTURED_1000000025_ATTORNEYS_EYES_ONLY_PS_STDNT_AWARDS" and "UCHICAGO_STRUCTURED_1000000026_ATTORNEYS_EYES_ONLY_PS_STDNT_AWD_PER_DATA_TABLE." UCHICAGO_STRUCTURED_1000000025 provides a field named "fin_aid_type" which has values "G", "S", "W", "L", "T", "V", and "" (missing). "G" and "S" refer to grants and scholarships, respectively, but these terms do not denote whether the award is need-based or merit-based. Therefore, one cannot distinguish need-based awards from merit-based awards using this field. Further, none of the variables in UCHICAGO_STRUCTURED_1000000025 allow one to expressly identify the source of the funding, which I require in order to distinguish institutional from non-institutional awards. After reviewing Dr. Hill's backup, I find that he uses the field "award_msg_cd" to distinguish institutional from other sources of awards. This is not a valid solution. Dr. Hill drops 38 percent of award observations that have a missing value for this field. The other dataset for 2016-2023 awards that Chicago produced was UCHICAGO_STRUCTURED_1000000026, which has fields specifying total institutional awards, total institutional need-based awards, total federal awards, and total federal need-based awards. The issue with these data is that they do not allow one to separate out institutional need-based *gift aid* awards from institutional need-based *loan* awards.

FIGURE 1: CHANGE IN AVERAGE REAL EFFECTIVE INSTITUTIONAL PRICES WHEN SWITCHING FROM CHICAGO'S PRE-2016 PFAIDS DATA TO 2016-2022 PEOPLESOFT DATA



Notes: I create this figure using Hill's backup production of Chicago's data.

112. These substantial, mismeasured increases in real Effective Institutional Prices, which are observed over the entire 2016-2023 period due to database inconsistencies, render Chicago's 2016-2023 data unreliable for use in my analyses.²⁴⁹ To see why, consider the time periods over which Chicago participated in the 568 Group. Chicago began to participate in the 568 Group starting in 2003, and then no longer participated from 2015 onwards. As shown in Figure 1, inconsistencies in Chicago's database systems result in real Effective Institutional Prices being improperly recorded by higher amounts from 2016-2023—exactly during those years that Chicago had ceased to

249. This discernible change in Chicago's real Effective Institutional Prices attributable to its change in database is different from what economists refer to as "classical measurement error." Rather, this measurement error is a systemic consequence of the change in database system. *See, e.g.,* Daniel L. Millimet and Christopher F. Parmeter, *Accounting for Skewed or One-Sided Measurement Error in the Dependent Variable*, IZA INSTITUTE OF LABOR ECONOMICS, Discussion Paper No. 12576 (Aug. 2019), 1 ("While classical measurement error in the dependent variable in a linear regression framework results only in a loss of precision, non-classical measurement error can lead to estimates which are biased and inference which lacks power.").

participate in the Challenged Conduct. In rudimentary terms, my regression model compares a Defendant's real Effective Institutional Prices during years in which the Defendant had participated in the 568 Group to years that they did not participate, as well as to other Defendants' real Effective Institutional Prices during years that those other Defendants did not participate. This means that the regression attributes the misreported higher real Effective Institutional Prices to periods that a Defendant did not participate in the Challenged Conduct.

113. As a result of this misattribution of higher real effective prices to non-participatory periods, the inclusion of the 2016-2023 Chicago data substantially biases my conduct coefficient downwards. To show this, I make a single adjustment to Dr. Hill's data provided in his backup by excluding Chicago's 2016-2023 data, and I run the same regression that he runs to generate his Figure 33.²⁵⁰ I find that the conduct coefficient from Dr. Hill's model 6 (which corresponds to my primary overcharge model) increases from \$441 to \$750 once I exclude Chicago's 2016-2023 data, a 70 percent increase in the conduct coefficient relative to what Dr. Hill obtains.

114. Dr. Hill makes another significant change to my processing of Duke's data that he does not note anywhere in his report, which I reject in computing my revised regressions. For context, Duke produced financial aid awards data that includes different award variables, such as variables for "offered" amounts, variables for "accepted" amounts, and variables for "paid" amounts. Upon review of these data, my team discovered that the total "paid" amount did not match the total "accept" amount for 87 percent of observations in these awards data, and that the total "accepted" amounts were over \$13,000 greater than the total "paid" amounts on average. Because the "paid" amounts

250. Dr. Hill's Figure 33 shows the cumulative effect of applying his data-processing adjustments, control-variable construction adjustments, and standard-error adjustments to my regression models. The comparison that I provide here is for the Model 6 of that figure, which corresponds to what I refer to as my primary overcharge model, which includes all of the control variables outlined in my Initial Report, as well as student-Defendant fixed effects. Dr. Hill reports the numeric values used in his Figure 33 in Appendix H.2 of his report (at Figure 79).

appeared incomplete, my team used the “accepted” amounts for purposes of measuring Duke’s financial aid. Dr. Hill alters my code instead to use “paid” amounts, and he provides no basis for doing so, nor does he explain anywhere in his report that he made this change.

115. This alteration to my code is wrong. To illustrate, I review named plaintiff Sia Henry’s financial aid records in Duke’s data, and I compare her total “paid” and “accepted” amounts to her actual financial aid awards that are referenced in Dr. Stiroh’s report and which come from Henry’s Supplemental Responses and Objections to Defendants’ First Set of Interrogatories. Dr. Stiroh notes that Sia Henry had been awarded \$9,300 for the 2007 academic year, \$14,200 for the 2008 academic year, \$40,455 for the 2009 academic year, and \$30,579 for the 2010 academic year.²⁵¹ In Duke’s financial aid data, I find exact matches to Sia Henry’s total aid awards when using the “accepted” award amounts, consistent with the “accepted” award amount being the *actual* award. The “paid” award amounts do not match for *any* year of Sia Henry’s data, and these “paid” award amounts vary substantially from the actual amounts Henry received. For instance, Sia Henry’s “paid” total award for 2007 academic year is \$0 (instead of \$9,300), for the 2008 academic year is \$0 (instead of \$14,200), for the 2009 academic year is \$34,800 (instead of \$40,455), and for the 2010 academic year is \$23,579 (instead of \$30,579).²⁵² I therefore find that Dr. Hill’s adjustment to use “paid” amounts instead of “accepted” amounts for Duke to be improper.

116. I adjust Dr. Hill’s code to use Duke’s accepted amount instead of their paid amount, and I find that correcting for this error further increases the conduct coefficient in Dr. Hill’s regressions. As noted above, making a single change to Dr. Hill’s methodology by excluding Chicago’s 2016-2023 data results in the conduct coefficient increasing from \$441 to \$750. Correcting

251. Stiroh Report ¶18. Plaintiff Sia Henry’s Supplemental Responses and Objections to Defendants’ First Set of Interrogatories, *Henry v. Brown University*, Case No. 1:22-cv-00125 (N.D. Ill. Jun. 27, 2023) at 4-5.

252. See my workpapers. Sia Henry can be identified in the structured data as UID_0000404135288.

Dr. Hill's code to use the proper financial aid award variables in Duke's data results in the conduct coefficient increasing from \$750 to \$930, a 24 percent increase in the conduct coefficient relative to when using the improper Duke awards variables.

117. Dr. Hill recodes the Challenged Conduct dummy variable for Rice and for Duke, which fundamentally alters the conduct variable and thereby changes the overcharge calculation. In my Initial Report, I defined Rice as having participated in the Challenged Conduct from 2003 to 2022. Dr. Hill adjusts my Class Period definition by coding Rice as having not participated in the Challenged Conduct from 2012 to 2014.²⁵³ He also adjusts my Class Period definition by coding Rice and Duke as having not participated in 2022, contrary to my definition, which assigned both Defendants as having participated during the 2022 academic year. While Dr. Hill includes this Class Period change in his section pertaining to "data processing" critiques, a change such as this is not what one would refer to as an error in "data processing." This is a legal issue as to how one is to interpret the record evidence. Counsel has instructed me to accept Dr. Hill's adjusted Class Period definition to be consistent with the record evidence that he cites.

118. Note that the changes that I make here only pertain to Dr. Hill's data-processing adjustments. In the next section, I make further corrections to Dr. Hill's adjustments pertaining to my control-variable construction, and I find that doing so further increases his conduct coefficient to a level similar to that which I showed in my Initial Report.

253. Treating Rice as having participated from 2012-2014 is conservative considering that there is record evidence that Rice left the 568 Group largely because it did not want to pay membership dues, not because of any ideological difference or limitations placed by the 568 Group. Rice continued to meet with 568 Group schools during COFHE and CNAR meetings during this period as well. *See, e.g.,* RICE_LIT0000007623; RICE_LIT0000007155; RICE_LIT0000033957; RICE_LIT0000067255.

ii. Dr. Hill Makes Unfounded Claims Regarding My Control-Variable Construction

119. In addition to the plethora of data-processing adjustments that Dr. Hill makes to my regression data, he also makes numerous *ad hoc* changes to the control variables that I use in my regression analysis. Dr. Hill does not criticize my decision to include the control variables that I chose. Instead, he levies his criticisms toward the form in which I included my controls, my treatment of potential outlying values, and my use of 1997 data. With respect to the last critique, regarding my use of the 1997 tuition revenue data, I accept his suggested correction. Upon incorporating it into my model, I find that it has no effect on my results. As for the rest of his adjustments that I discuss in turn below, his criticisms are misplaced, and his claimed “corrections” violate commonly accepted statistical principles.

(a) Treatment of Potential Outlying Values

120. Dr. Hill asserts that my student-level control variables for family-adjusted gross income and net worth have extreme outlier values, and he claims: “It is well known in the economic and statistics literature that including outliers in regressions can affect the estimation results substantially.”²⁵⁴ To deal with these purported outlier values, Dr. Hill applies a method known as winsorization. This technique involves replacing values of a variable that lie at the upper and lower bounds of its distribution. In other words, winsorization replaces more extreme data (data that lie at the tails of a variable’s distribution) with less extreme data (i.e., data that lie closer to the mean). Dr. Hill winsorizes family-adjusted gross income and net worth at the 1st and 99th percentiles—that is, for each of these variables, Dr. Hill replaces values that are less than the 1st percentile with the 1st percentile value, and replaces values that are greater than the 99th percentile with the 99th percentile

254. Hill Report ¶196.

value. Dr. Hill's decision to make these adjustments rests upon the assumption that values beyond the 1st and 99th percentiles represent "outliers."

121. Dr. Hill's methodology suffers from critical flaws: (1) he misinterprets the term outlier and conflates "outliers" with "influential observations" (2) he does not follow the prescription for outlier detection in his own cites; and (3) he mechanically modifies data without acknowledging patterns in the data that militate against his arbitrary winsorization.

122. Before delving into the errors inherent in Dr. Hill's decision to winsorize the data, I emphasize that the "outlier" issue that Dr. Hill raises pertains to only two out of the twelve control variables in my primary regression model, without even counting the 285,476 student dummy variables that are controlled for through the use of student fixed effects.²⁵⁵ Even if these two control variables were subject to a small number of outliers, this feature would be irrelevant to my findings.

123. To begin, Dr. Hill neither defines the term "outlier" nor explains why the top and bottom one percent of values automatically qualify as "outliers." Generally, outliers are observations that "often appear at either end of the cumulative distribution."²⁵⁶ Dr. Wooldridge, whose textbook Dr. Hill cites in various sections of his report but not with respect to outliers, explains outliers as follows:

Outliers can also arise when sampling from a small population if one or several members of the population are very different in some relevant aspect from the rest of the population. The decision to keep or drop such observations in a regression analysis can be a difficult one, and the statistical properties of the resulting estimators are complicated. Outlying observations can provide important information by increasing the variation in the explanatory variables (which reduces standard errors).²⁵⁷

255. See Singer Report Table 11, Column 6. My primary regression model that I use for purposes of computing aggregate damages includes student fixed effects. I also show the same regression model including only school fixed effects in columns 1-3 of Singer Report Table 11.

256. DAVID BELSLEY, EDWIN KUH, AND ROY WELCH, REGRESSION DIAGNOSTICS: IDENTIFYING INFLUENTIAL DATA AND SOURCES OF COLLINEARITY 19 (Wiley Interscience 2004) [hereafter BELSLEY, KUH, AND WELCH (2004)].

257. See WOOLDRIDGE at 327.

The econometric literature draws a distinction between outliers and influential observations. Influential observations exert “influence” on the regression line by modifying regression coefficients:

Data sets sometimes contain a small number of observations that are separated from the rest of the data, in the sense that they have values that deviate strongly from the other observations. Potentially, such ‘extreme cases’ can become influential observations that affect the results of the regression analysis. An influential observation can be defined as a case ‘that alters the value of a regression coefficient whenever it is deleted from an analysis.’²⁵⁸

124. Further, “outliers and leverage points are not necessarily influential cases.”²⁵⁹ As a result, Dr. Hill’s use of winsorization to account for outliers is questionable because it is overbroad and incorrectly converts a large amount of data without reasonable justification. Moreover, research has found that “winsorizing does little to mitigate the effect of influential observations.”²⁶⁰ In other words, to the extent that Dr. Hill wants to minimize the effects of influential observations, (1) he conflates outliers and influential observations, and (2) winsorizing does not accomplish this goal. Further, Belsley, Kuh and Welch also point out, “Unusual or influential data points, of course, are not necessarily bad data points,”²⁶¹ and researchers should not reflexively delete or modify them. Dr. Hill does not investigate whether any of these data points are true outliers, leverage points, or influential observations. He simply posits an *ad hoc* rule by which he changes the actual data to a value closer to the mean, without investigating the data points. Renowned statistician John Tukey warned against the sort of *ad hoc* approach that Dr. Hill takes, explaining that, upon encountering extreme values in the tails of a distribution:

[W]e are likely to think of them as ‘strays’ [or] ‘wild shots’ . . . and to focus our attention on how normally distributed the rest of the distribution appears to be. One who does this commits two oversights, forgetting Winsor’s principle that ‘all distributions are normal in the middle,’

258. BART MEULEMAN, GEERT LOOSVELDT, AND VIKTOR EMONDS, REGRESSION ANALYSIS: ASSUMPTIONS AND DIAGNOSTICS, IN SAGE HANDBOOK OF REGRESSION AND CAUSAL INFERENCE 102 (Henning Best and Christof Wolf eds. 2015) [hereafter SAGE HANDBOOK OF REGRESSION].

259. *Id.*

260. Andrew J. Leone, Miguel Minutti-Meza, and Charles E. Wasley, *Influential Observations and Inference in Accounting Research*, 94(6) THE ACCOUNTING REVIEW 337–364 (2019).

261. BELSLEY, KUH, AND WELCH (2004) at 3.

and forgetting that *the distribution relevant to statistical practice is that of the values actually provided and not of the values which ought to have been provided.*²⁶²

The National Bureau of Standards (now National Institute for Standards and Technology, or NIST) has proffered the same cautionary note.²⁶³ Tukey also adds that: “Sets of observations which have been de-tailed by over-vigorous use of a rule for rejecting outliers are inappropriate, since they are not samples.”²⁶⁴ Dr. Hill offered no indication that any of the values I used were data-entry errors or were not relied upon by Defendants when making financial aid decisions. Nor did he provide any statistical analysis that would support his treatment of two percent of the data as outliers. Finally, he offered no justification for his *ad hoc* application of winsorization. As I explain above, his decision to winsorize departs from commonly accepted statistical practice.

125. Next, while Dr. Hill cites Choi (2009), a political science paper, as ostensibly supporting his outlier argument, Choi explains “I employ two different outlier diagnostics: a partial regression plot and DFBETAs.”²⁶⁵ In contrast, Dr. Hill performs no such analysis, choosing instead just to assume that the top and bottom one percent of the data (two percent total) merit modification. Statisticians and econometricians recognize that not all outliers represent influential observations and that only influential observations merit investigation.

262. John W. Tukey, *A Survey of Sampling from Contaminated Distributions*, in *Contributions to probability and statistics* in ESSAYS IN HONOR OF HAROLD HOTELLING 448-485, 457 (Ingram Olkin et al. eds. Stanford Univ. Press 1960) [hereafter Tukey (1960)].

263. P.E. Pontius, *Measurement Philosophy of the Pilot Program for Mass Calibration*, U.S. DEPARTMENT OF COMMERCE, NATIONAL BUREAU OF STANDARDS, Technical Note 288 (May 6, 1966) at 5-6, <https://nvlpubs.nist.gov/nistpubs/Legacy/TN/nbstechnicalnote288.pdf> (“A major difficulty in the application of statistical methods to the analysis of measurement data is that of obtaining suitable collections of data. The problem is more often associated with conscious, or perhaps unconscious, attempts to make a particular process perform as one would like it to perform rather than accepting the actual performance...Rejection of data on the basis of arbitrary performance limits severely distorts the estimate of the real process variability...Realistic performance parameters require the acceptance of all data that cannot be rejected for cause.”).

264. Tukey (1960) at 458.

265. Seung-Whan Choi, *The effect of outliers on regression analysis: regime type and foreign direct investment*, 4(2) QUARTERLY JOURNAL OF POLITICAL SCIENCE 153-65, 155 (2009).

126. Researchers have several tools at their disposal to investigate such potentially influential observations.²⁶⁶ Dr. Hill employs none of them, simply choosing to arbitrarily apply one percent winsorizing. Moreover, I expect Dr. Hill would have had access to individuals at Defendant schools who could have provided more insight into data points that might be considered outliers. He makes no mention of having discussed such issues to support his treatment of potential outliers. Knowing whether such data points represent true values resulting from the data-generating process is critical. The only example that Dr. Hill provides to support his assertion that his purported “outliers” are in fact outliers is that a single student, UID_0000122606635, has a recorded adjusted student gross income of -99999999.²⁶⁷ This single example does not provide a basis for modifying the top and bottom one percent of the *entire* dataset—which alters the values of adjusted gross income and net worth for over 24,000 Class Member-academic year combinations in total.²⁶⁸

127. Dr. Hill’s winsorization suffers from yet another important flaw. The purported “outlier” values that Dr. Hill replaces are not random but are instead correlated by academic year and by Defendant. To see this, consider Figure 2 below, which plots the percent of observations that Dr. Hill replaces over time by winsorizing at the 1st and 99th percentile values. Dr. Hill replaces an increasing proportion of adjusted gross income and net worth observations over time, as shown by the upwards trend in both the “% AGI Changed” and “% NW Changed” lines. His methodology replaces a significantly higher proportion of net worth observations from 2009–2013 relative to other

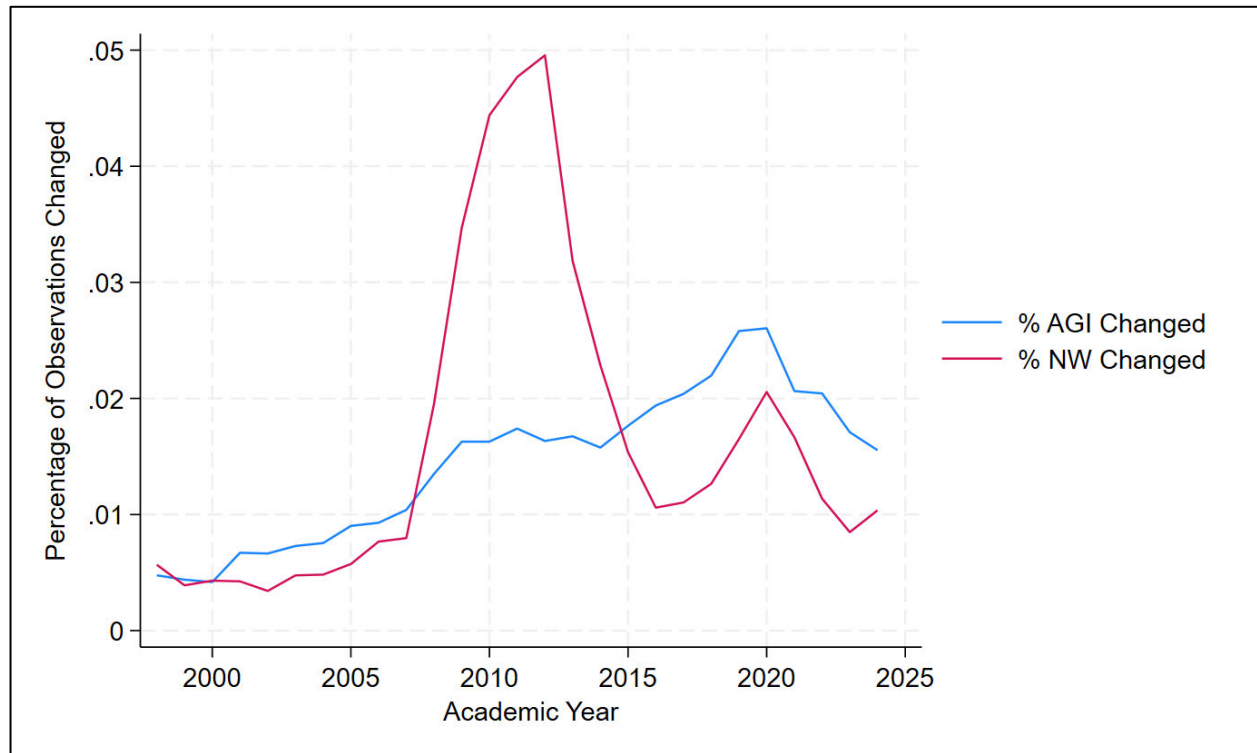
266. SAGE HANDBOOK OF REGRESSION at 102 (“Influential observations can be identified by evaluating the impact that single cases have on the regression outcomes. We will discuss three commonly used measures to quantify this impact, namely the DFFITS, Cook’s distance and DFBETA.”).

267. Hill Report n. 291.

268. See my workpapers.

years. This spike in the proportion of net worth observations replaced is consistent with the time frame of the Great Recession, which resulted in a significant decrease in U.S. household net worth.²⁶⁹

FIGURE 2: PERCENT OF OBSERVATIONS OVER TIME THAT DR. HILL REPLACES USING HIS WINSORIZATION METHOD

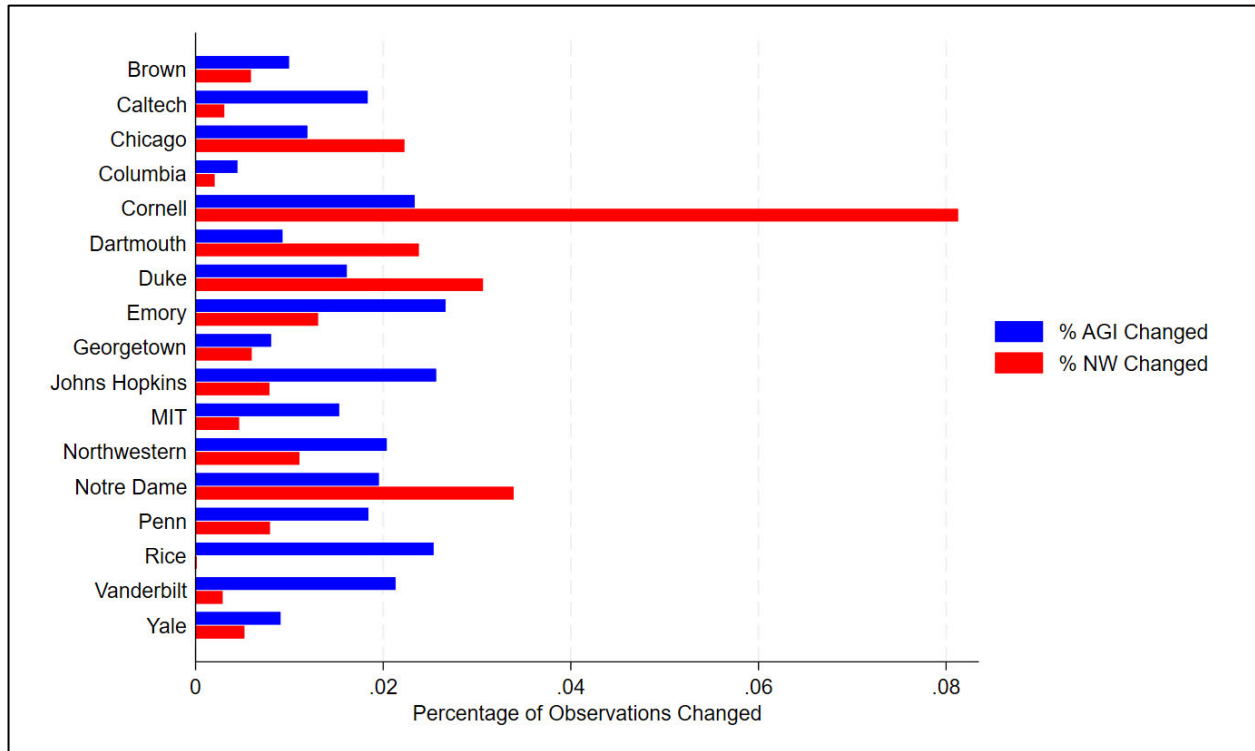


Notes: “AGI” is adjusted gross income. “NW” is net worth.

128. Figure 3 plots the proportion of observations by Defendant that Dr. Hill replaces via his winsORIZATION. The proportions vary substantially across Defendants. A significant share of Cornell net worth observations have their net worth replaced relative to other Defendants. Chicago, Dartmouth, Duke, and Notre Dame also have much larger proportions of net worth observations replaced compared to Caltech, Columbia, Rice, and Vanderbilt.

269. *How Recessions Impact Household Net Worth*, FEDERAL RESERVE BANK OF ST. LOUIS (Nov. 23, 2020), <https://www.stlouisfed.org/on-the-economy/2020/november/recessions-impact-household-net-worth/> (“This recession was more than twice as long as the previous two (18 months) and had the greatest impact on the bottom 50% of households, costing them as much as 42% of their net worth during the downturn, Mendez-Carbajo found.”).

FIGURE 3: PERCENT OF OBSERVATIONS BY DEFENDANT THAT DR. HILL REPLACES USING HIS WINSORIZATION METHOD



Notes: “AGI” is adjusted gross income. “NW” is net worth.

129. The heterogeneity in Dr. Hill’s winsORIZATION across time and across Defendants produces unreliable results. Any correlation between the values that Dr. Hill winsORIZES and the conduct variable can introduce substantial bias by artificially altering the relationship between Effective Institutional Prices and the Challenged Conduct.²⁷⁰ For instance, consider Yale, which ended its participation in the 568 Group in 2008 and rejoined in 2018. Yale’s conduct variable therefore switches from one to zero concurrent to the increase in Dr. Hill’s replacement of net worth values during the Great Recession. Brown and Emory both participated in the 568 Group from 2004 to 2012, resulting in their conduct variable values being correlated with Dr. Hill’s net worth

270. Joe H. Sullivan, Merrill Warkentin, and Linda Wallace, *So many ways for assessing outliers: What really works and does it matter?*, 132 JOURNAL OF BUSINESS RESEARCH 530-543, 535 (2021) (“Winsorizing represents somewhat of a compromise between ignoring extreme cases and purging them. However, there is no guidance from theory as to the fraction to use, it can make the sample less representative, and it is a univariate technique, not considering correlations.”). *See also id.* at 540 (“For example, replacing an outlier, which is common in Winsorizing, can lead to substantial bias.”).

winsorization. Additionally, Cornell, which has the highest proportion of net worth observations replaced, had participated in the Challenged Conduct for almost all years for which they produced data. In contrast, Brown has a significantly lower proportion of net worth observations replaced, yet produced data containing far more years in which Brown had *not* participated in the Challenged Conduct compared to the number of years for which Brown *did* participate.

130. I evaluate the artificial downward effect on the conduct coefficient introduced by Dr. Hill's winsorization of adjusted gross income and net worth. I find that the conduct coefficient in Dr. Hill's version of my primary overcharge model is \$1,108 after accepting his proper data-processing adjustments, and removing his improper data-processing adjustments.²⁷¹ When I correct Dr. Hill's code by preserving his purported outliers as-is, apart from the one clear outlier example that he proffered for student UID_0000122606635, which I drop from the data, I find that Dr. Hill's overcharge estimate increases from \$1,108 to \$1,218—a 10 percent increase.²⁷²

(b) Extrapolation for the 1997 Academic Year

131. Dr. Hill critiques my use of an extrapolation for the academic year 1997 for the institutional tuition revenue control variable.²⁷³ This extrapolation of 1997 values comes from the fact that I use a lagged version of this variable. Put differently, the value I assign to 1998 is meant to reflect the 1997 value. Because I align all data according to Defendants' data timeline of 1998 through the present, I set 1998 as the starting point for institutional tuition revenues and extrapolated the 1997 values as a result of this truncation. I accept Dr. Hill's adjustment to include the 1997 institutional tuition revenue values; it has no material impact on my results.²⁷⁴

271. This is after correcting Dr. Hill's data to: (1) exclude Chicago's post-2015 data, and (2) use Duke's "accepted" amounts instead of their "paid" amounts.

272. Both coefficients are statistically significant when using my robust standard errors.

273. Hill Report ¶197.

274. Dr. Hill's adjustment of my lagged tuition revenue control variable slightly *increases* my primary model conduct coefficient from \$1,216 to \$1,218.

(c) Dr. Hill's Improper Categorization of Continuous Variables

132. Dr. Hill claims that I was inconsistent in my coding of a student's year in college, and he implies that my implementation of the student's year in college and number of family members in college were improper because I applied this variable in its actual continuous form.²⁷⁵ While I accept Dr. Hill's minor adjustments to how I had coded the year in college variable, I disagree with his erroneous attempt to categorize these two continuous variables, year in college and number of family members.

133. Dr. Hill contends that instead of using the actual number of years in college or number of family members in my regression, I should have instead created dummy variables for each possible value. For instance, the student's year of college variable could be split into five dummy variables. The first dummy variable would equal to one in observations where a student is a first year and zero otherwise, the second dummy variable would equal to one in observations where a student is a second-year and zero otherwise, and so on. Put differently, Dr. Hill suggests that these variables should enter the model as separate, binary dummy variables for each value.

134. Transforming a continuous variable into multiple variables, one for each value or group of values in the continuous variable, is known by various names, such as "categorization," "binning," "discretization," and, in cases where only two categories are created, "dichotomization." While Dr. Hill calls his approach more "flexible," economists and statisticians have long recognized the perils of categorizing continuous variables, as doing so entails losing data, or information.²⁷⁶

275. Hill Report ¶199.

276. Douglas Altman and Patrick Royston, *The Cost of Dichotomising Continuous Variables*, 332 THE BMJ (2006) ("Dichotomising leads to several problems. Firstly, much information is lost, so the statistical power to detect a relation between the variable and patient outcome is reduced. Indeed, dichotomising a variable at the median reduces power by the same amount as would discarding a third of the data.").

Another renowned statistician, Jacob Cohen, advised that using all the information in a continuous variable is superior to dichotomization.²⁷⁷

135. Altman (2014) explains that “Categorization of continuous data is not necessary, and indeed is not a natural way of analyzing continuous data for most statisticians.”²⁷⁸ Instead of this dummy variable approach, I enter these variables in their original continuous, linear form. As Cartensen (2020) explains, “categorization assumes that the relationship between the predictor and the response is flat within intervals; this assumption is far less reasonable than a linearity assumption in most cases.”²⁷⁹ My linear application of these variables follows from a simple economic logic: the more family members that a student has who attend a postsecondary educational institution, the greater the student’s need and thus the greater the student’s financial aid. Likewise, as a student proceeds through their education they are less likely to depart and thus the institution can offer less aid than in their first year. My linear application of these control variables is therefore reasonable, and I find Dr. Hill’s adjustment of these numerical variables to be unnecessary and contrary to statistical principles discussed in the relevant literature.

(d) I Use a Commonly Accepted Approach to Adjust for Zero Values When Using Logs

136. Hill claims that in my log-linear models, which I provided in Appendix 4 Tables 2-3 of my Initial Report, I improperly constructed certain control variables.²⁸⁰ Specifically, I had added

277. Jacob Cohen, *The Cost of Dichotomization*, 7(3) APPLIED PSYCHOLOGICAL MEASUREMENT 249-253, 252 (1983) (“It is, of course, possible to construct non-normal bivariate distributions where dichotomization results in an increase in r over that of the original graduated variables; but these will be characterized by extreme skewness, heteroscedasticity, and curvilinearity, e.g., step functions. Needless to say, such distributions are uncommon in the behavioral and social sciences. When they do occur, dichotomization is a far inferior approach to one that uses all the measurement information in the original graduated data and tackles the curvilinearity directly, e.g., polynomial regression.”).

278. DOUGLAS ALTMAN, *CATEGORIZING CONTINUOUS VARIABLES* (John Wiley & Sons 2005).

279. Bendix Carstensen, *Do Not Group Quantitative Variables*, EPIDEMIOLOGY WITH R (Oxford University Press 2020).

280. Hill Report ¶198.

\$1,000 to income, \$1,000 net worth, and \$1 to non-institutional grant aid before converting these control variables to logarithms. He incorrectly states that this adjustment is unreliable.

137. My modification of these three control variables is standard practice in economics as a method to handle negative and zero values in a log-linear model.²⁸¹ The natural logarithm function does not exist for numbers less than or equal to zero, because no number exists that, when raised to any power, equals zero or a negative number. Because researchers frequently apply the natural log functional form to variables that contain zero values, options for accounting for such values include (a) converting values that are less than or equal to zero to values marginally greater than zero, (b) dropping observations with values less than or equal to zero, or (c) functional form transformations such as inverse hyperbolic sine.

138. I chose to convert values less than or equal to zero using the common method of adding a sufficiently large enough constant to render the values all marginally greater than zero. My approach is well accepted in the economic literature.²⁸² Dr. Hill instead adds a dummy variable for each of these variables that equals to one whenever the variable equates to a negative or zero value. This methodology is not standard. The footnote that Dr. Hill cites to in support of this method is a reference regarding the purpose of using a dummy variable *generally*, it does not have anything to do

281. Thomas E. MaCurdy & John H. Pencavel, *Testing between Competing Models of Wage and Employment Determination in Unionized Markets*, 94(3) JOURNAL OF POLITICAL ECONOMY S3-S39 (Jun. 1986). *See also* Marc Bellemare, *Metrics Monday: What to Do Instead of $\log(x + 1)$* , MARCFBELLEMARE.COM (Feb. 26, 2018), <https://marcfbellemare.com/wordpress/12856> (“When you want to log a variable x but that x has many zero-valued observations, there are three things you can do in principle:... Use $\log(x + 1)$, $\log(x + 0.001)$, or some variant thereof. This is a method that MaCurdy and Pencavel introduced in a 1986 JPE article, and it has long been the workhorse way to deal with those wayward zero-valued observations.”) IBM, in user guide for its SPSS software, also recommends the same approach. *See also Can you perform a log transformation in SPSS?*, IBM SUPPORT (Apr. 16, 2020), <https://www.ibm.com/support/pages/can-you-perform-log-transformation-spss> (“Can you perform a log transformation in SPSS? . . . If there are cases with values of 0 for X , you will need to add a constant to X before taking the log, as the log of 0 is undefined. You can add a constant of 1 to X for the transformation, without affecting X values in the data, by using the expression $\ln(X+1)$.”). *See also* WOOLDRIDGE at 193 (“In cases where a variable y is nonnegative but can take on the value 0, $\log(1+y)$ is sometimes used. The percentage change interpretations are often closely preserved, except for changes beginning at $y = 0$ (where the percentage change is not even defined). Generally, using $\log(1+y)$ and then interpreting the estimates as if the variable were $\log(y)$ is acceptable when the data on y contain relatively few zeros.”).

282. *See* n. 281, *supra*.

with using dummy variables for this specific purpose. Rather, Dr. Hill's adjustment suppresses all variation when these variables are equal to or less than zero. For example, Dr. Hill's regression treats a student with zero net worth equivalent to a student whose family is burdened by substantial debt such that their net worth falls negative. My modification accounts for this variation and is therefore more reliable.

iii. Dr. Hill's Assertion That My Regressions Produce Counter-Intuitive Findings Reflects His Misinterpretation of Causal Inference

139. Dr. Hill claims that the coefficients and standard errors on my control variables imply results that are inconsistent with economic theory.²⁸³ For instance, he points out that my coefficients on adjusted gross income and net worth are lower in magnitude than one might expect, and that my coefficients for other grant aid and lagged excess endowment return are not directionally consistent with economic theory.²⁸⁴

140. The econometric literature cautions against ascribing a causal interpretation to control variables, as Dr. Hill erroneously attempts to do.²⁸⁵ Hünermund & Louw (2023) explain the practice as follows:

Control variables are included in regression analyses to estimate the causal effect of a treatment on an outcome. [...] we recommend refraining from interpreting the marginal effects of controls and focusing on the main variables of interest, for which a plausible identification argument can be established. To prevent erroneous managerial or policy implications, *coefficients of control variables should be clearly marked as not having a causal interpretation* or omitted from regression tables altogether. Moreover, *we advise against using control variable estimates for subsequent theory building and meta-analyses*.²⁸⁶

141. My inclusion of other variables in my regressions is consistent with the literature. I include other variables to control for confounders—variables that can create “back-door paths”

283. *Id.* ¶200.

284. *Id.*

285. Paul Hünermund and Beyers Louw, *On the Nuisance of Control Variables in Causal Regression Analysis*, 0 ORGANIZATIONAL RESEARCH METHODS (2023).

286. *Id.* (emphasis added).

between the treatment variable (the Challenged Conduct) and the outcome of interest (Effective Institutional Prices), and that thus inhibit the recovery of a causal effect from the treatment.²⁸⁷ The field of causal inference, which has widespread cross-disciplinary application in fields including economics, epidemiology, and others, distinguishes between types of explanatory variables (“right-hand side variables” in a regression). Dr. Hill ascribes a causal interpretation to my control variables, which is commonly referred to as the “Table 2 Fallacy.”²⁸⁸ Ascribing a causal interpretation to all coefficients in a model presumes all are mutually adjusted—that is, they are all controls for each other, and, by including these variables in a model, they close any back-door paths between each other and the outcome. This is at best unlikely and almost certainly not true. The control variables in my model to which Dr. Hill erroneously ascribes causal interpretations—adjusted gross income, net worth, other grant aid, and lagged excess endowment returns—are not *themselves* controlled for by other variables in the model. Rather, these four variables that Dr. Hill interprets are included to isolate the effect of the Challenged Conduct on Effective Institutional Prices by controlling for any confounding factors that might influence Effective Institutional Prices. Because these four variables bear a plausible relationship to Effective Institutional Prices, which Dr. Hill does not contest, including them in my regression is consistent with standard statistical practice. Other than serving to

287. For a succinct definition of confounders, see Judea Pearl, *On a Class of Bias-Amplifying Variables that Endanger Effect Estimates*, PROCEEDINGS ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE (UAI2010), Technical Report R-356 (Jul. 2010), https://ftp.cs.ucla.edu/pub/stat_ser/r356.pdf (“The common method of reducing confounding bias in the analysis of causal effects is to adjust for a set of variables judged to be “confounders,” that is, variables capable of producing spurious associations between treatment and outcome, not attributable to their causal dependence. It is well known that a sufficient condition for the elimination of confounding bias is that the set of adjusted variable be “admissible,” namely, that it satisfies the back-door criterion.”).

288. Daniel Westreich and Sander Greenland, *The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients*, 177(4) AMERICAN JOURNAL OF EPIDEMIOLOGY 292-298 (2013). See also *The Table 2 Fallacy*, DAGGITY.NET, <https://dagitty.net/learn/graphs/table2-fallacy.html> (last visited Oct. 2024) (“One particularly widespread misconception is known as mutual adjustment, recently called the Table 2 fallacy since the first table in most epidemiological articles usually describes the study data, and the second table reports the results of a multivariable regression model where the erroneous efforts to illustrate mutual adjustment often appear.”).

close potential back doors between the treatment and outcome, these control variables carry no other interpretation.

142. Aside from the irrelevance of Dr. Hill's review of my specific control-variable results for the reasons stated above, I also find his arguments unconvincing. For instance, Dr. Hill notes that my adjusted gross income and net worth control variables are correct directionally—that is, both variables' coefficients are positive, consistent with theory.²⁸⁹ Dr. Hill's critique is simply that these two control variables' coefficients are not high enough. He insinuates that these variables' coefficients should instead show a “strong” or “significant” relationship with Effective Institutional Prices, but he does not opine on what the magnitude of a “strong” or “significant” coefficient would be.²⁹⁰ Similarly, Dr. Hill notes that my coefficient for other grant aid, which is equal to the sum of a student's non-institutional grant aid, is *not always* directionally consistent with economic theory. Yet this coefficient on other grant aid for my primary model is directionally consistent with economic theory, contrary to his argument.

**iv. Dr. Hill Exaggerates the Import of His Data-Processing
“Corrections”**

143. After Dr. Hill makes his plethora of data-processing and control-variable-construction changes listed in Table 5 above, he then reruns my regressions and evaluates their output. He alleges that, as a result of his changes, the fit of the regression improves substantially, the coefficients on the control variables better align with economic theory, and that my overcharge model no longer produce any reliable evidence of overcharges.²⁹¹ Dr. Hill's model is flawed, as is his evaluation of its “reliability.”

289. Hill Report ¶200.

290. *Id.* Dr. Hill also critiques that my coefficient on adjusted gross income is not statistically significant. As explained in Part III.A.2.b, Dr. Hill generally elevates statistical significance to an undue level of importance, inconsistent with the ASA Position Statement.

291. *Id.* ¶201.

144. First and foremost, Dr. Hill's only presents the effect of his data-processing and control-variable changes after also replacing my robust standard-error computations with his own two-way clustered standard errors. He claims that my models show "no reliable evidence of overcharges" after he makes these changes.²⁹² This is incorrect. My primary model continues to show a positive overcharge if I were to accept *every* change that Dr. Hill's proffers, as shown by the blue bar in model 6 of his Figure 33.²⁹³ Dr. Hill conflates a lack of statistical significance when using his improper clustered standard errors to a zero overcharge, but he ignores the fact that the coefficient point estimates are positive and economically significant. Importantly, as expounded in Part III.A.2.b above, applying Dr. Hill's clustered standard errors to my model results in artificially inflated *p*-values via an artificial decrease in sample size. Therefore, that he finds no statistical significance in my primary model is invalid. If I alter Dr. Hill's methodology to instead apply my robust standard-error methodology to his regressions and his data, I obtain statistically significant results for every regression shown in his Figure 33.²⁹⁴

145. Second, I take issue with many of Dr. Hill's data-processing and control-variable-construction adjustments, and when I account for this fact, his findings change substantially. As I explained above, Dr. Hill erroneously adds Chicago financial aid data for the period 2016-2023. These data are imprecisely measured relative to the pre-2016 Chicago data, and adding Chicago financial data for this period significantly biases his conduct coefficient estimates downwards. Moreover, Dr. Hill uses the wrong Duke financial aid award variables, and this biases his conduct coefficient estimates downwards further. Dr. Hill also improperly modifies data based on what he

292. *Id.* §8.2.4.

293. Dr. Hill obtains a conduct coefficient of 441.1 using my primary regression model after he makes all adjustments, which corresponds to an artificial inflated in Effective Institutional Prices of \$441.1 per Class Member and academic year. *See* Hill Report Figure 79.

294. *See* my workpapers for details.

claims to be “outliers,” and he improperly enters my year in college and number in college control variables into the model categorically rather than continuously. When I take account of all these errors and combine them with those adjustments proffered by Dr. Hill that I do accept, I find, as I discuss in more detail below, that my revised primary overcharge regression continues to show that the Challenged Conduct had a statistically significant impact in artificially inflating Effective Institutional Prices.

146. Dr. Hill also makes several statistically unsound arguments in support of why his models are superior to my models. Dr. Hill first opines that his models result in a “better fit” to the data compared to my models because they produce higher R-squared values.²⁹⁵ This is in tension with the econometrics and statistical literature, which advises against using R-squared values for evaluating the reliability of a statistical model.²⁹⁶ Further, because Dr. Hill converts the continuous variables for year in college and number in college into separate dummy variables for each value, this mechanically must increase the R-squared because R-squared increases whenever an additional variable is added to a regression.²⁹⁷ Nonetheless it is worth noting that, for my original primary regression model, I obtained an R-squared value of 84 percent, and for my updated primary regression model that includes the adjustments that I accept and describe above, I obtain an R-squared value of 87 percent. Dr. Hill obtains an R-squared of 85 percent for his version of my primary regression model, which is lower than for my updated model.²⁹⁸

295. *Id.* ¶202.

296. *See, e.g.*, WOOLDRIDGE at 39 (“Students who are first learning econometrics tend to put too much weight on the size of the R-squared in evaluating regression equations. For now, be aware that using R-squared as the main gauge of success for an econometric analysis can lead to trouble.”); JAMES H. STOCK AND MARK W. WATSON, INTRODUCTION TO ECONOMETRICS 237 (Addison Wesley 2nd ed. 2006) at 238 [hereafter STOCK & WATSON (2006)] (“The R^2 and \overline{R}^2 do NOT tell you whether: 1. An included variable is statistically significant; 2. The regressors are a true cause of the movements in the dependent variable; 3. There is omitted variable bias; or 4. You have chosen the most appropriate set of regressors.”).

297. *See, e.g.*, STOCK & WATSON (2006) (“The R^2 increases whenever you add a regressor, whether or not it is statistically significant.”).

298. Hill Report Figure 79.

147. Dr. Hill goes on to claim that his model provides a better “fit” to the data because the coefficients on his control variables have the expected signs. This is misleading. Adjusted gross income, net worth, and other gift aid all have positive coefficients in my Initial Report primary regression, as stated above.²⁹⁹ Furthermore, as explained above in Part III.A.2.c.iii, the economic and statistical literature advises against assigning a causal interpretation to control variables, contrary to Dr. Hill’s justification for his model “fit.”

148. Dr. Hill asserts that my observation counts and R-squared values for my student fixed effects regressions are “misleading” because my models do not use any student singleton observations for identification.³⁰⁰ Put differently, because my student fixed effects models only consider variation across time within each student, any student with a single observation does not affect the regression coefficients. This is not standard practice. It is commonplace in the economic literature to report summary regression statistics based on the underlying data, and to not make these unorthodox adjustments after regressing.³⁰¹ I maintain consistency in my reporting of observation counts and R-squared values between both my Defendant fixed effects regressions (in Table 11 columns 1-3 of my Initial Report) and student fixed effects models (in Table 11 columns 4-6 of my Initial Report). Further, I use a common software that is widely used in the economics profession—STATA—and I report the statistics as it outputs them. This is not “misleading.”

299. Singer Report Table 11 Column 6. The only control variable that Dr. Hill mentions as being directionally inconsistent with expectations is lagged excess endowment investment returns, which has a positive coefficient in all of my Table 11 regressions.

300. *Id.* n. 318.

301. *See, e.g.,* Reply by Carlo Lazzaro, *Fixed Effects Regression and the role of Singleton Observations in a Balanced Panel*, STATALIST (Jan. 31, 2023 at 10:08), <https://www.statalist.org/forums/forum/general-stata-discussion/general/1656468-fixed-effects-regression-and-the-role-of-singleton-observations-in-a-balanced-panel> (“report data as they are in descriptive statistic tables (that is, including singletons the community-contributed module - reghdfe- omits.”).

v. My Updated Regressions Continue to Show a Positive and Statistically Significant Effect of the Challenged Conduct on Effective Institutional Prices Paid by Class Members

149. As explained above, I accept a number of Dr. Hill's adjustments to my data processing, and I accept a few adjustments that he makes to my control variable construction. Each adjustment that I accept is outlined in Table 5 above with a value of "Y" in the "Accept?" column. As described in Part III.A.2.c.i, many of these adjustments were the result of contradictions between Defendants' data and/or responses compared to Dr. Hill's interpretation of these data, or due to Defendants not having produced sufficient data prior to my Initial Report submission. The adjustments that Dr. Hill makes and that I accept are either due to (1) a clarification from Defendants as to the interpretation of their data; (2) additional data that should be incorporated into my analyses; or (3) corrections to minor errors in my code.

150. I find that the cumulative effect of making these adjustments is modest. In Table 6, I provide my updated regression results.³⁰² The updated conduct coefficient in my primary model in

302. In my workpapers, I provide revised versions of the Effective Institutional Price and institutional grant aid regression tables I provided in my Initial Report, which use the revised data adjusted to incorporate all of Dr. Hill's data processing adjustments marked as "Accepted" = "Y" outlined in Table 5. I note that columns 1-2 of my institutional grant aid regression tables no longer consistently show a negative conduct using these updated data. I also note that columns 1-3 of my Effective Institutional Price regression tables using a logarithmic functional form no longer show positive conduct coefficients. These results only pertain to my Defendant-only fixed effects regressions, but not my Defendant-student fixed effects regressions, which still produce conduct coefficients consistent with the Challenged Conduct resulting in an artificial overcharge in Effective Institutional Prices and an artificial underpayment in institutional grant aid. As I explained in my Initial Report, my primary model includes student fixed effects because these account for time-invariant, student-specific factors that might otherwise confound the relationship between the dependent variable and the conduct variable. *See* Singer Report ¶245. These results are consistent with such a concern and support my decision to rely on a specification using Defendant-student fixed effects as my primary model for determining common impact and damages. I also explain this point in Part III.A.2.g, *infra*. I also specifically use my column 6 regression, as opposed to columns 4-5 (which also include student fixed effects), because the column 6 regression not only includes student-Defendant fixed effects, but also incorporates all of the student, institution, and macroeconomic control variables detailed in my Initial Report, and thereby is most robust to any potential omitted variable bias. It also bears noting that my institutional grant aid and logarithmic Effective Institutional Price regressions are distinct from my primary model.

column 6 shows that a Defendant's participation in the Challenged Conduct results in a \$1,202 artificial increase to their Effective Institutional Prices. My conduct coefficient decreases in light of the myriad adjustments that Dr. Hill proffers and that I accept, from 1,497 to 1,202; the conduct coefficient remains positive and statistically significant at a one percent significance level across all six specifications of my regression. These results are consistent with the Challenged Conduct having artificially inflated Class Members' Effective Institutional Prices relative to a but-for world absent the Challenged Conduct.³⁰³

Additionally, my Effective Institutional Price regression using a levels specification more closely matches the data-generating process than the logarithmic specification. The formulaic approach that Defendants use in their needs analysis reflects a levels relationship between independent variables and the EFC. For example, an additional family member in college generally reduces a student's FM EFC by a dollar amount, not a percentage. *See, e.g., The EFC Formula, 2014-2015*, U.S. DEPARTMENT OF EDUCATION, FEDERAL STUDENT AID, <https://fsapartners.ed.gov/sites/default/files/attachments/efcformulaguide/091913EFCFormulaGuide1415.pdf> (last visited Oct. 2024) (showing an additional \$4,820 income protection allowance for a family of four versus a family of five, with one in college); *id.* at 10 (showing a student's income assessed at 0.5 and assets assessed at 0.2).

303. In my reply workpapers, I provide the other alternative regressions that I had included in the appendices of my Initial Report. Alternative versions of my primary regression model show that the Challenged Conduct resulted in an artificial inflation in Effective Institutional Prices and an artificial suppression of institutional grant aid.

-104-

TABLE 6: UPDATED EFFECTIVE INSTITUTIONAL PRICE REGRESSION RESULTS

	Dependent Variable: <i>Real Effective Institutional Price</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
	<i>No Student Fixed Effects</i>			<i>Includes Student Fixed Effects</i>		
Conduct	523.21***	524.46***	945.17***	996.44***	1,241.21***	1,201.94***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Adj. Gross Income	54.54***	54.37***	54.35***	31.00***	30.84***	30.67***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Net Worth	1.61***	1.59***	1.58***	0.30***	0.29***	0.28***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Number in College	-2,845.42***	-2,817.33***	-2,823.26***	-5,782.52***	-5,770.70***	-5,763.23***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Year in College	582.08***	551.29***	558.47***	1,483.39***	1,313.33***	-987.41***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Student's Gift Aid from Other Sources	-0.08***	-0.08***	-0.08***	0.33***	0.33***	0.32***
	(0.002)	(0.003)	(0.003)	(0.000)	(0.000)	(0.000)
Lagged Excess Endowment Investment Returns		462.88***	-262.17**		1,096.76***	1,083.26***
		(0.000)	(0.038)		(0.000)	(0.000)
Inst. Tuition Rev per FTE Undergraduate (1-year lag)		0.05***	0.02***		0.11***	0.05***
		(0.000)	(0.000)		(0.000)	(0.000)
% of FY-FTE Undergrads Receiving Financial Aid		-89.69***	-78.41***		-24.59***	-32.89***
		(0.000)	(0.000)		(0.000)	(0.000)
Unemployment (1-year lag)			-197.44***			-120.30***
			(0.000)			(0.000)
COVID			4,644.11***			3,358.89***
			(0.000)			(0.000)
Trend			-152.23***			2,128.17***
			(0.000)			(0.000)
Real GDP			0.47***			0.54***
			(0.000)			(0.000)
Observations	701,220	701,220	701,220	701,220	701,220	701,220
R-Squared	0.19	0.19	0.19	0.87	0.87	0.87
Includes Institution Fixed Effects?	Y	Y	Y	Y	Y	Y
Includes Institution*Student Fixed Effects?	N	N	N	Y	Y	Y
Number of Fixed Effects	17	17	17	257,769	257,769	257,769

Notes: Robust p-values in parentheses; ***p<0.01, **p<0.05, *p<0.1. Adjusted Gross Income and Net Worth reported in thousands. These regressions are equivalent to the regressions that I provided in Table 11 of my Initial Report, but using revised data adjusted to incorporate all of Dr. Hill's data-processing adjustments marked as "Accepted" = "Y" outlined in Table 5.

d. I Employ Standard Econometric Methods to Measure the Effect of the Challenged Conduct on Effective Institutional Prices on a Classwide Basis

151. In Section 8.3 of his rebuttal report, Dr. Hill levies various misplaced criticisms regarding my choices of control variables to include in my econometric model. His arguments are irrelevant at and in direct conflict with the econometric literature. I address each in turn.

i. Dr. Hill Incorrectly Defines Omitted Variable Bias, Contradicting the Literature and Undermining His Critiques

152. Dr. Hill's critiques regarding my control variables focus on his claim that my results suffer from omitted variable bias. His arguments are inapposite, most importantly because *Dr. Hill*

incorrectly defines omitted variable bias. Had Dr. Hill relied on an actual econometrics text instead of an incorrect definition that appears in the American Bar Association’s “Econometrics” book, he would have noticed the incongruity between his definition of “omitted variable bias” and the definition that appears in the econometrics literature.

153. Dr. Hill incorrectly defines omitted variable bias:

...[Dr. Singer’s] model does not adequately control for the changing cost of higher education and the lasting impact of shocks such as the COVID-19 pandemic and subsequent, rapid inflationary pressures. These factors may be unrelated to the Challenged Conduct, but if they are not adequately controlled for, they might be captured by Dr. Singer’s estimated impact of the Challenged Conduct, rendering his results unreliable. Without appropriate controls the estimated effect of the conduct may suffer from omitted variable bias and may not provide a reliable estimate of overcharges.³⁰⁴

Dr. Hill rebuts his own argument of omitted variable bias. By definition, omitted variable bias cannot occur if the omitted variable is “unrelated to the Challenged Conduct.” If the variables that were omitted are not related to the conduct, then such variables cannot bias my conduct coefficient. The bias occurs only to the extent that the variables that I excluded bear some relation to the conduct (as well as to the dependent variable).³⁰⁵ If the omitted variables are independent of the conduct, then no bias occurs. Adding to the model independent variables that are uncorrelated to the conduct might serve to improve the fit of the model, but they do not address any omitted variable bias, because excluding such variable would not have biased the conduct coefficient in the first place.

304. Hill Report ¶206.

305. WOOLDRIDGE at 90 (“Thus, we have the important conclusion that, if x_1 and x_2 are uncorrelated in the sample, then $\hat{\beta}_1$ is unbiased.”) *See also* SAGE HANDBOOK ON REGRESSION 68 (“If important independent variables which are both related to the dependent and the independent variables are left out of the model the estimates are biased – a bias also referred to as *omitted variable bias*.”) *See also* Hal Singer and Kevin Caves, *Applied Econometrics: When Can an Omitted Variable Invalidate a Regression*, 17(3) *The Antitrust Source* (2017) (“Intuitively, omitted variable bias occurs when the independent variable (the X) that we have included in our model picks up the effect of some other variable that we have omitted from the model. The reason for the bias is that we are attributing effects to X that should be attributed to the omitted variable.”). *See also* R. Wilms, E. Mäthner, L. Winnen, and R. Lanwehr, *Omitted variable bias: A threat to estimating causal relationships*, 5 *METHODS IN PSYCHOLOGY* 1-10 (2021) (“We would like to create a common understanding about what constitutes the omitted variable bias. In short, *the omitted variable bias emerges if an omitted third variable causes the independent and dependent variable*.”) (emphasis added).

154. Dr. Hill bases his inaccurate definition of omitted variable bias on an American Bar Association text titled “Econometrics: Legal, Practical, and Technical Issues.” While this text may be informative generally on practical applications of econometrics in the legal field, it is not the authoritative text on econometric issues. Econometric texts are legion, and Dr. Hill could have consulted virtually any of them, such as the Wooldridge text I cite above and throughout this report, and that he also cites elsewhere in his report for other reasons, for an accurate definition of omitted variable bias.

155. To make matters worse, Dr. Hill then cites to the same ABA text to argue that a regression model should include *all relevant variables*. He cites the following excerpt from ABA’s Econometrics:

The parameter estimate on the conspiracy period indicator variable could then be used to assess the overcharge that occurred as a result of the conspiracy. For this approach to yield reliable estimates of the overcharge, the model must include all factors that are likely to affect price. Otherwise, the parameter estimate on the conspiracy indicator variable may be confounded with other explanatory variables omitted from the model.³⁰⁶

Reliance on his citation above leads Dr. Hill astray on two fronts. *First*, as I explained, an excluded variable can cause omitted variable bias only to the extent that it is related to the treatment variable—in this case, the Challenged Conduct (i.e., the conspiracy period indicator). If, as Dr. Hill suggests, no such correlation exists, then no omitted variable bias exists. *Second*, the claim in the ABA’s Econometrics text that a model “must include all factors that are likely to affect price” diverges not only from the literature, but also from legal interpretation thereof. While I have no opinion on legal matters, I understand that courts have acknowledged this reality, holding that a regression need not include all relevant variables.³⁰⁷

306. Hill Report n. 322.

307. For a discussion of the pertinent cases dealing with the inclusion of relevant variables in a regression analysis, with an application to antitrust matters, see Ted Tatos, *Relevant Market Definition and Multi-Sided Platforms Post Ohio v. American Express: Evidence from Recent NCAA Antitrust Litigation*, 10(2) HARVARD JOURNAL OF SPORTS AND ENTERTAINMENT LAW 147-172 (2019).

156. Based on the tenor of his critiques, Dr. Hill appears to stake out the position that if the coefficient of the variable of interest (i.e., the treatment variable, in this case, the conduct coefficient) changes when excluding certain variables, then such excluded variables must be confounders that belong in the regression. To the extent that he adopts this position, he is wrong.

157. The causal-inference literature recognizes that variables that appear on the right-hand side of the regression (aka, independent variables) can adopt different roles: for example, these can serve the role of confounders, proxy confounders, or mediators. Mediators are variables that lie in the causal pathway between the treatment and the outcome; in other words, they provide an indirect path through which the treatment can affect the outcome. In this case, they can act as “bad controls.” As Angrist and Pischke explain:

We’ve made the point that control for covariates can increase the likelihood that the regression estimates have a causal interpretation. *But more control is not always better. Some variables are bad controls and should not be included in a regression model even when their inclusion might be expected to change the short regression coefficients.* Bad controls are variables that are themselves outcome variables in the notional experiment at hand. That is, bad controls might just as well be dependent variables too. Good controls are variables that we can think of having been fixed at the time the regressor of interest was determined.³⁰⁸

Angrist and Pischke explain that timing matters. For example, if increasing higher-education costs motivated Defendants to engage in the Challenged Conduct, then such costs would provide an alternative avenue for the effect of the Challenged Conduct on the prices Class Members paid Defendants. Moreover, if the COVID-19 pandemic, or aspects thereof, allowed Defendants any additional latitude to suppress institutional aid and thus raise prices, then the pandemic also serves as a conduit rather than confounding factor. In other words, such variables can be mediators, not confounders.

308. JOSHUA D. ANGRIST AND JÖRN-STEFFEN PISCHKE, *MOSTLY HARMLESS ECONOMETRICS: AN EMPIRICIST’S COMPANION* 64 (Princeton University Press 2009) (emphasis added). *See also* Carlos Cinelli, Andrew Forney, and Judea Pearl, *A Crash Course in Good and Bad Controls*, 53(3) JOURNAL SOCIOLOGICAL METHODS AND RESEARCH 1071-1104 (2024).

158. To the extent that I included such variables as control factors in my model, then my conduct coefficient would underestimate the effect of the Challenged Conduct. By controlling for a mediator, I would have closed off an indirect avenue through which the Challenged Conduct could have suppressed financial aid. My conduct coefficient would then only capture the direct effect of the Challenged Conduct rather than the total effect, which would include indirect avenues. Given that I did not have perfect information, as is often the case in empirical studies, I chose to adopt a conservative approach. Nevertheless, the bottom line is that my analysis showed that the Challenged Conduct resulted in economically and statistically significant artificial inflation in Effective Institutional Prices.

ii. Dr. Hill Incorrectly Defines and Misinterprets the Concept of “Statistical Significance”

159. Dr. Hill almost exclusively focuses his critiques of both my report and that of Dr. Bulman on changes Dr. Hill imposes that affect the statistical significance of my results. For example, Dr. Hill’s report contains 188 instances of the phrase “statistically significant” and 11 instances of the phrase “statistical significance.” In contrast, terms such as “economically significant”, “practically significant”, and “materially significant” never appear in his report.

160. Two critical flaws plague his focus on statistical significance. First, Dr. Hill incorrectly defines statistical significance. He offers this incorrect explanation:

A coefficient that is statistically significant at the 1 percent level has 1 chance in 100 of being a false positive, a coefficient that is statistically significant at the 5 percent level has 1 chance in 20 of being a false positive, and a coefficient that is statistically significant at the 10 percent means that the 10 percent cutoff is ten times more likely to accept a false positive than is the 1 percent cutoff and five times more likely to accept a false positive than is the 5 percent cutoff.

Dr. Hill’s attempted definition of statistical significance is incorrect. Indeed, he does not cite any statistical authority as support. He again appeals to the American Bar Association’s text, “Proving

Antitrust Damages,” a text he cites throughout his report.³⁰⁹ Dr. Hill also references the Reference Manual on Scientific Evidence, Reference Guide on Statistics, but (1) his citation offers only a general description of statistical testing with no mention of the p-values that Dr. Hill uses, and (2) as I explain below, he erroneously conflates p-values and error rates, two incongruous statistical concepts.

161. Throughout his report, Dr. Hill uses p-values as the basis for his determination of statistical significance, the same approach that I used in my report. A staple of Fisherian³¹⁰ statistics, the term p-value refers to the probability of obtaining a result (defined as the difference between observed and expected) as large or larger, assuming that the null hypothesis is true.³¹¹ In other words, a null hypothesis posits that the conduct coefficient is zero, then asks, if that were true, what is the probability of obtaining a t-score (standardized effect, defined as the coefficient divided by its standard error) as large or larger. For ease of interpretation, practitioners have used certain cutoffs to signal whether the p-value is “statistically significant.” I agree with Dr. Hill that no “bright line” rule exists for determining statistical significance. As I note below, the American Statistical Association makes this same point.

162. Yet p-values cannot and should not be interpreted as false positive rates, as Dr. Hill does. Statisticians Greenland et al., whose paper reviewing misinterpretations of p-values merited special acknowledgment by the American Statistical Association in preparing the ASA statement on p-values, explain:

If you reject the test hypothesis because $P \leq 0.05$, the chance you are in error (the chance your “significant finding” is a false positive) is 5 %. No! To see why this description is false,

309. Hill Report n. 386, citing AMERICAN BAR ASSOCIATION, PROVING ANTITRUST DAMAGES: LEGAL AND ECONOMIC ISSUES 143-44 (ABA Book Publishing 3rd ed. 2017).

310. I refer to Fisherian statistics as the methodology adopted by the renowned statistician Ronald Aylmer (R.A.) Fisher.

311. See also ASA Position Statement (“Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data (e.g., the sample mean difference between two compared groups) would be equal to or more extreme than its observed value.”).

suppose the test hypothesis is in fact true. Then, if you reject it, the chance you are in error is 100%, not 5 %. The 5% refers only to how often you would reject it, and therefore be in error, over very many uses of the test across different studies when the test hypothesis and all other assumptions used for the test are true. It does not refer to your single use of the test, which may have been thrown off by assumption violations as well as random errors.³¹²

163. Moreover, statistical significance reflects an interpretation of the p-value, not a property of the effect studied (i.e., the Challenged Conduct). Statistical significance represents the dichotomization of the p-value into “significant” or “not significant.” Dr. Hill erroneously interprets this as the property of the population. He claims that a coefficient significant at the five percent level (i.e., $p \leq 0.05$) indicates a false positive chance of five percent. His interpretation is wrong. If a p-value is 0.024, then the probability of finding a result as large or large if the null is true equals 2.4 percent, not 5 percent. Moreover, p-values have no long-run implications, as such they are not error rates. Dr. Hill confuses two different statistical principles, undermining his conclusions.

164. Second, Dr. Hill claims:

If the coefficient for an independent variable (e.g., lagged excess returns) is not statistically significant, its correlation with the dependent variable (e.g., net price) is not statistically distinguishable from zero and there is no basis for concluding that there is a meaningful relationship between them.³¹³

Dr. Hill’s position is categorically incorrect. The American Statistical Association again cautions against this sort of erroneous logic. In its position statement on p-values, the ASA explains:

Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold. Practices that reduce data analysis or scientific inference to mechanical “bright-line” rules (such as “ $p < 0.05$ ”) for justifying scientific claims or conclusions can lead to erroneous beliefs and poor decision making. A conclusion does not immediately become “true” on one side of the divide and “false” on the other. The widespread use of “statistical significance” (generally interpreted as “ $p < 0.05$ ”) as a license for making a claim of a scientific finding (or implied truth) leads to considerable distortion of the scientific process.³¹⁴

312. Sander Greenland, Stephen J. Senn, et al., *Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations*, 31 EUROPEAN JOURNAL OF EPIDEMIOLOGY 337-350, 342 (2016) (emphasis added). See also Ronald L. Wasserstein and Nicole A. Lazar, *The ASA’s Statement on p-Values: Context, Process, and Purpose*, 70(2) THE AMERICAN STATISTICIAN 129-33 (2016).

313. Hill Report ¶241.

314. *Id.* at 131 (emphasis added).

Wooldridge's Introductory Econometrics text, which Dr. Hill cites throughout his report, also contradicts Dr. Hill on his singular focus on statistical significance.³¹⁵ Dr. Hill runs afoul of commonly accepted statistical practice when he elevates statistical significance to the critical level of decision making as the ASA explained above.

165. As explained in my Initial Report, I did not base my opinions on statistical evidence alone, contrary to the approaches Defendants' Experts took. Consistent with the ASA's guidance on this matter, which represents the accepted statistical practice as reflected in the literature, I leveraged not only the data but contextual factors and record evidence in supporting my opinions. In my Initial Report, I repeatedly explain that my results are both statistically and economically significant, consistent with ASA guidelines. As such, I reject Dr. Hill's critiques, which both misinterpret and misapply statistical significance.

166. Moreover, when interpreting his own results, Dr. Hill's interpretation of results changes. When my results show a negative sign on the conduct variable, he insists that his preferred statistical significance threshold must also be met. But when he interprets the results of his own affirmative analysis, which he claims shows a positive effect on grant aid from the conduct, he explains his results as: "A positive sign indicates the alleged conduct increased financial aid,"³¹⁶ even though none of his results are statistically significant.

315. See Singer Report n. 372, citing WOOLDRIDGE at 135-6 ("Because we have emphasized *statistical significance* throughout this section, now is a good time to remember that we should pay attention to the magnitude of the *coefficient* estimates in addition to the size of the *t* statistics. The statistical significance of a variable x_j is determined entirely by the size of $t_{\hat{\beta}_j}$, whereas the *economic significance or practical significance* of a variable is related to the size (and sign) of $\hat{\beta}_j$. [] Too much focus on statistical significance can lead to the false conclusion that a variable is "important" for explaining y even though its estimated effect is modest.") (emphasis added).

316. Hill Report Figure 12. Moreover, despite the fact that he criticizes my use of the CPI, he also deflates his own results by the CPI, just as I did.

iii. Dr. Stiroh Repeats Dr. Hill's Error and Likewise Misinterprets Statistical Results

167. Like Dr. Hill, Dr. Stiroh confuses statistical and practical significance. For example, instead of using a common conduct coefficient across Defendants, to reflect the collusive nature of the Challenged Conduct, she uses different conduct coefficients for each Defendant. As I explained in Section III.A.2.a.i of this report, the common coefficient that I use reflects the appropriate method of recovering the causal effect of the Challenged Conduct. In addition to inappropriately estimating separate effects by Defendant, Dr. Stiroh selectively interprets her results.

168. When her results show no statistically significant overcharge (she makes no mention of the practical or economic significance), she concludes:

For six of the Defendants, Dr. Singer's model yields no statistically significant overcharge in the specification that Dr. Singer uses for his damages calculation. Therefore, even Dr. Singer's model does not show an overcharge for Class members that attended these Defendant schools.³¹⁷

Based solely on lack of statistical significance, Dr. Stiroh erroneously concludes that such a result is evidence of no overcharge. Contrary to well-accepted statistical and economic principles, including the guidance from the American Statistical Association, Dr. Stiroh gives economic, or practical, significance, no weight. Her argument reflects the position that statistical significance is the sine qua non of analytical evidence. As I explained above, even if I agreed with her decision to estimate separate coefficient for each Defendant, which I do not, she draws the incorrect conclusion from her results.

iv. My Regressions Properly Account for Cost Changes Over Time

169. Dr. Hill argues that my econometric models should have controlled for costs, even though I accounted for changes in costs by performing all calculations in real dollars (i.e., adjusting

317. Stiroh Report ¶11(v)(b)(1).

for inflation by converting nominal values into real terms using the CPI). My doing so reflects a standard methodology employed by researchers, as explained in my original report. Dr. Hill cites no academic papers that include additional costs beyond adjusting for the CPI when performing a similar analysis to mine. Instead, Dr. Hill claims:

Changes in costs are reflected in prices and hence need to be accounted for when trying to analyze changes in prices over time. For example, an increase in the salaries paid to university professors will lead to higher costs for universities and likely higher tuition and higher Effective Institutional Prices. If a regression attempting to explain Effective Institutional Prices does not control for such changes in costs, those changes in cost could result in omitted variable bias as it is instead captured by other variables (e.g., a conduct variable).³¹⁸

170. To begin, Dr. Hill cites to no authority to defend his position. His cite to Davies and Garces produces no support for his broad claim. The authors only contemplate running a regression of prices on exogenous variables “and our instrumental variable—perhaps cost data.”³¹⁹ He paints with too broad a brush when attempting to deduce that a general claim that costs affect prices means that I should have also included other cost variables in my analysis in addition to deflating by the CPI. He offers no paper specific to higher education to support his position. In fact, the relevant literature militates *against* his argument.

171. Specifically, Dr. Hill offers the higher education price index (HEPI) as an alternative to the CPI and suggests that this variable belongs in my regression equation. Yet he never attempts to include HEPI. Instead, he simply muses that its exclusion may cause some omitted variable bias. But, as explained earlier, if the HEPI is not related to the Challenged Conduct but instead only explains additional variation in price, then it cannot cause omitted variable bias. Dr. Hill’s argument here again reflects his erroneous definition of statistical concepts.

318. *Id.* ¶210.

319. PETER DAVIS AND ELIANA GARCÉS, QUANTITATIVE TECHNIQUES FOR COMPETITION AND ANTITRUST ANALYSIS 104 (Princeton University Press 2010).

172. Moreover, in their 2011 paper titled *Stop Misusing Higher Education Specific Price Indices*, Gillen and Robe (2011) caution against the type of logic that Dr. Hill adopts. Focusing specifically on the HEPI, the authors explain:

There are a number of reasons to suspect that HEPI and HECA do not give an accurate measure of the change in the cost of providing an education because they both suffer from several systematic biases...The first issue with quality stems from the fact that HEPI and HECA are input-based price indices. Unlike most price indices (such as the CPI or the GDP IPD), input-based indices do not directly measure the price of the good or service in question. Instead, they measure changes in the cost of providing it. This would be akin to the CPI measuring changes in the price of apples, not by directly measuring the price of apples, but rather by measuring the price of the seeds, fertilizer, labor, and transportation needed to grow and sell apples.³²⁰

173. The authors delineate several sources of bias that plague the HEPI: quality bias, productivity bias, and substitution bias, noting that HEPI is also self-referential. The latter point particularly exposes the error in Dr. Hill's argument. Dr. Hill posits, without support, that increasing professor salaries will result in "likely higher tuition and higher Effective Institutional Prices."³²¹ As an economic matter, I concur that some institutions might pass along higher professional-compensation costs in the form of higher tuition. But this issue is peripheral to the central argument. The key question is whether (1) Defendant schools passed along increases in professional compensation to students and (2) whether changes in that pass-through related to the Challenged Conduct. Dr. Hill's argument fails to answer either of these questions, let alone both of them. Gillen and Rose point out the logical flaw in Dr. Hill's speculation:

HEPI is self-referential in the sense that it relies upon labor costs (faculty, administrative, and clerical salaries) which are influenced by university policy decisions. Richard Vedder further observed that this is a problem because if administrative salaries rise, then the Higher Education Price Index rises. Colleges can give their employees huge salary increases, claim

320. Andrew Gillen and Jonathan Robe, *Stop Misusing Higher Education Specific Price Indices*, CENTER FOR COLLEGE AFFORDABILITY AND PRODUCTIVITY (2011) [hereafter "Gillen and Robe"].

321. Hill Report ¶210.

that ‘higher education costs are soaring,’ and demand larger government subsidies, etc., as a consequence.³²²

Consistent with my use of CPI, Gillen and Robe conclude that when adjusting tuition for inflation, the proper approach is to use the CPI.³²³ As such, I reject Dr. Hill’s unsupported claims that I should have used an alternative measure. Moreover, I already include a time trend variable. As Dr. Hill’s own Figure 34 shows, because HEPI follows an increasing trend, the trend I include already captures this effect. To wit, the time trend variable shows a 99.4 percent correlation with HEPI over the 1999-2022 period.

174. Finally, Dr. Hill’s argument ignores the basics of higher education. As Bowen’s Revenue Theory of Cost explains, schools have the incentive to spend all their revenues.³²⁴ Dr. Hill’s cost justification would allow schools to suppress financial aid but spend those savings on new buildings or other prestige-enhancing features. Dr. Hill would then argue that the latter justifies the former.

v. Dr. Hill’s Suggested Alternative Specifications for Time Trend and COVID-19 Controls Are Inapposite and Self-Contradictory

175. In my econometric model, I incorporated various controls for economic factors, including the annual Gross Domestic Product, a variable to capture the 2020 effects of the COVID-19 pandemic, and a time trend to capture additional time-related economic factors. Dr. Hill claims that my controls for numerous economy-wide shocks are too restrictive, and that (1) instead of using

322. Gillen and Robe at 12, citing Richard K. Vedder, *Going Broke by Degree: Why College Costs Too Much*, WASHINGTON DC: AMERICAN ENTERPRISE INSTITUTE (2004), at 41 and 5 Richard Vedder, *Federal Tax Policy Regarding Universities: Endowment and Beyond*, WASHINGTON DC: CENTER FOR COLLEGE AFFORDABILITY AND PRODUCTIVITY (2008).

323. Gillen and Robe at Table 2 and 14 (“For most purposes (including indexing college tuition for inflation), the use of HEPI or HECA is simply inappropriate.”).

324. Robert B. Archibald & David H. Feldman, *Explaining Increases in Higher Education Costs*, 79(3) THE JOURNAL OF HIGHER EDUCATION 268–295, 269 (2008) (“In Howard Bowen’s view, the source of cost increases in higher education is the rising revenue stream made available to colleges and universities. Higher education institutions spend everything they can raise, so revenue is the only constraint on cost.”).

a time trend, I should have used a dummy variable for each year, and (2) instead of using a dummy variable to capture the effects of COVID-19 in the 2020-21 academic year (2020 data year), I should have included a dummy variable for each year affected by COVID-19, from 2020 through the end of the data.

176. By juxtaposing his proposed corrections, the critical flaw in his argument becomes immediately apparent. If one includes dummy variables for each year to capture the time trend, as he suggests, then one cannot also include dummy variables for 2020-2024 to separately capture the effects of COVID-19. Doing so would result in two sets of dummy variables for 2020-2024. Dr. Hill fails to consider the combined effects of his supposed “corrections.”

177. The 2020 academic year comprised of fall 2020 through spring 2021. The most prominent effects of COVID-19 on Effective Institutional Prices would occur during that year.³²⁵

325. This is evident from Defendants’ statements and policies when setting tuition rates for the 2020-2021 and 2021-2022 academic years. In 2020, the University of Chicago released a statement stating that “In response to these challenging times, there will be no increase in the combined total of tuition, housing, and fees for College students in the 2020-2021 academic year.” (emphasis omitted). Ka Yee C. Lee & John W. Boyer, *2020-2021 College Tuition, Housing and Fees*, THE UNIVERSITY OF CHICAGO (Apr. 13, 2020), <https://college.uchicago.edu/2020-2021-college-tuition-housing-and-fees>. In 2021 the University of Chicago had resumed tuition increases. Calculation of change in tuition (59,256 - 57,642 = 1,614). See *Costs*, THE UNIVERSITY OF CHICAGO <https://web.archive.org/web/20210301122831/https://financialaid.uchicago.edu/undergraduate/costs> (archived Mar. 1, 2021); *Costs*, THE UNIVERSITY OF CHICAGO, <https://web.archive.org/web/20220120155943/https://financialaid.uchicago.edu/undergraduate/costs> (archived Jan. 20 2022). The pandemic also prompted Duke University to rescind planned tuition increases for the 2020-2021 academic year. Planned tuition increases then resumed for the 2021-2022 academic year. See *Duke to Rescind Planned Undergraduate Tuition Increase, Reduce Fees for 2020-21 Academic Year*, DUKE TODAY (Aug. 1, 2020), <https://today.duke.edu/2020/08/duke-rescind-planned-undergraduate-tuition-increase-reduce-fees-2020-21-academic-year>; *Duke Trustees Set Tuition, Reappoint President at Quarterly Meeting*, DUKE TODAY (Feb. 27, 2021), <https://today.duke.edu/2021/02/duke-trustees-set-tuition-reappoint-president-quarterly-meeting>. Northwestern University offered a 10 percent tuition reduction for first- and second-year students for the Fall quarter of 2020 because of the COVID-19 pandemic, and then raised tuition by 3.5 percent in 2021. See Emma Edmund, *Northwestern goes remote for first- and second-year students, reduces Fall Quarter undergraduate tuition by 10 percent*, THE DAILY NORTHWESTERN (Aug. 28, 2020), <https://dailynorthwestern.com/2020/08/28/campus/northwestern-goes-remote-for-first-and-second-year-students-reduces-fall-quarter-undergraduate-tuition-by-10-percent/>; Yunkyo Kim, *Northwestern to increase total cost by 3.6 percent, financial aid by more than 8 percent*, THE DAILY NORTHWESTERN (Jun. 14, 2021),

Additionally, I find no basis to assume that lasting COVID-19 effects would be correlated with Defendants ending the conspiracy starting in 2023, considering the time gap between 2020 and 2023. Therefore, even if one were not to account for COVID-19 in the model (and, as stated, I do control for it), it would be unlikely to bias the conduct coefficient.

178. With respect to time dummies, Dr. Hill previously suggested that I should have included a control for the HEPI. If doing so were sufficient, then Dr. Hill's suggestion that I should also include year dummies appears superfluous. As noted above, HEPI has a 99.4 percent correlation with the time trend. Thus, including HEPI and the time dummies would have the same effect as including both a linear time trend and time dummies. Doing so would be improper, as the two would be perfectly collinear.

179. Further, time fixed effects are problematic here because many Defendants provided data that contain only one year to two years in which the Defendant did not participate in the conspiracy (i.e., 2023 and 2024). Time fixed effects mean including dummy variable for each year. Because the data I use are annual, the dummy variable for 2023 would absorb most of the conduct effect due to its high collinearity with the conduct variable. The result would significantly inhibit the model's ability to identify the effect of the conduct on schools that were only part of the conspiracy in 2023.

<https://dailynorthwestern.com/2021/06/14/campus/northwestern-to-increase-total-cost-by-3-6-percent-financial-aid-by-more-than-8-percent/>. The reduced tuition for the 2020-2021 academic year sometimes persisted into the 2021-2022 academic year, but national trends indicated that tuition raises resumed in following academic years. *See* Jennifer Ma and Matea Pender, *Trends in College Pricing and Student Aid 2023*, COLLEGE BOARD (2023), <https://research.collegeboard.org/media/pdf/Trends%20Report%202023%20Updated.pdf> (“tuition increases for the 2020-21 and 2021-22 academic years were among the lowest in recent decades. More colleges and universities raised tuition in 2022-23 and 2023-24 than in the previous two years.”).

e. Dr. Hill's Claim That I Should Have Omitted Post-Conduct Data Is Fatally Flawed

180. Dr. Hill argues that because Defendants provided incomplete data for the 2024 academic year, relying on the data that Defendants did provide renders my estimates “unreliable.”³²⁶ Dr. Hill suggests that because I did not have all of the data for 2024, I should have ignored the data that I did have. Dr. Hill also claims that data covering the 2023 academic year are incomplete—specifically, he shows that the number of students and total awards decline significantly in these years.³²⁷ To the extent that he argues I should have ignored these data, his argument is wrong.

181. As an initial matter, I understand that all Defendants were required to produce 2023-2024 data in a timely fashion. Dr. Hill's argument reflects a misleading attempt to deflect the responsibility from the Defendants to me. I used the data that Defendants provided. Cornell, Georgetown, JHU, and Penn produced financial aid awards data, and I understand that these data are incomplete and preliminary.³²⁸ Had I instead ignored these data, I suspect Dr. Hill would criticize me for not including them.

182. Dr. Hill offers no support for his claim to the effect that limited data should be ignored if the researcher cannot obtain complete data. Indeed, the application of this sort of logic would inhibit scientific research itself. Seldom do academic researchers and practitioners find themselves with perfect data. Defendants carry the responsibility of producing their data, and any such limitations reflect the longstanding nature of the alleged conspiracy and the fact that the 568 Cartel did not disband until recently. Defendants thus produced both limited pre-cartel data, given that the cartel began in 2003, and now argue that I cannot use portions of the limited post-cartel data.

326. *Id.*

327. *Id.*

328. *Id.* ¶225.

183. I reject such arguments as reflecting only attempts to erect artificial barriers to performing my analysis. Defendants' nearly two-decade alleged conspiracy has imposed data limitations, rendering any benchmark data particularly valuable. If the alleged conspiracy had lasted only one decade instead of nearly two, I would have had more benchmark data at my disposal. In this case, I used the full complement of benchmark data available to me.

184. I note that the same response applies to Dr. Hill's implication that extrapolating IPEDS data for 2023-2024 was somehow unwarranted.³²⁹ The 2023 data carry particular significance for my analysis; given the longstanding nature of the conspiracy, 2023 is the only year of data outside of the 568 agreement for many Defendants. Thus, to maximize the value of the available data, I employed commonly accepted statistical methods.

185. When dealing with a missing control variable data, one must either (1) extrapolate the control variable for the period of the missing data; (2) drop the control variable; or (3) drop the observations where the control variable is missing. Hill's solution is (3). The first option is the most reasonable approach here, given the significance of data during 2023 and given the relevance of including the control variable to address any potential omitted variable bias.

186. In my Initial Report, I stated that Effective Institutional Prices likely remained at least partially inflated after Defendants ended the 568 Group in 2022.³³⁰ I cited literature showing that there tends to be lingering effects of a cartel after it ceases to exist. Because of these lingering effects, my benchmark "clean" period after 2022 is likely partially contaminated by the Challenged Conduct, and this contamination would therefore render my overcharge estimate conservative. Dr. Hill asserts

329. *Id.*

330. Singer Report ¶228.

that my argument is invalid because when he drops data from the post-conduct period from my regressions, my overcharge coefficient becomes either zero or negative.³³¹

187. Dr. Hill confuses my argument with the idea that including post-conduct data should decrease the conduct coefficient. I argue that Effective Institutional Prices during the 2023 and 2024 academic years are likely still at least partially inflated, and so my conduct coefficient will be less than the true overcharge. One still must include the 2023-2024 data in order to properly measure the effect of the Challenged Conduct on prices because, as stated already, these data are relevant for a number of Defendants that produced limited “clean” data.

f. Dr. Stiroh’s Critique Regarding My Conduct Period Definition is Irrelevant

188. As explained in my Initial Report, my regressions exploit variation in Effective Institutional Prices between periods when Defendants had formally participated in the 568 Group to periods when they did not participate.³³² Dr. Stiroh points out that these participation periods are inconsistent with the periods specified in the Complaint.³³³ She states that “Plaintiffs allege that Defendants who formally departed from the 568 Group did not, in fact, cease the Challenged Conduct at those times and that the conduct continued until the present.”³³⁴

189. Dr. Stiroh’s critique is irrelevant to my opinions. I understand that Plaintiffs do not accept claims of withdrawal as a legal matter. I do not opine on whether Defendants’ claimed withdrawals, as a matter of antitrust conspiracy law, were effective. I relied upon dates upon which Counsel provided to me.

331. *Id.* ¶226.

332. Singer Report ¶227. *See also id.* at Appendix 3 Table 1 (showing years that each Defendant participated).

333. Stiroh Report ¶189.

334. *Id.*

g. Dr. Hill Is Incorrect in Asserting That My Student Fixed Effects Regression Models Are “Conceptually Flawed”

190. I presented two alternative sets of regressions in my Initial Report. For the first set of regressions shown in columns 1-3 of Table 11 in my Initial Report, I used Defendant fixed effects, whereas in the second set of regressions shown in columns 4-6, I used student and Defendant fixed effects. Fixed effects are a common, statistical methodology used throughout the economics literature to control for unobservable, time-invariant factors that affect the dependent variable in a regression.³³⁵ Student fixed effects can be interpreted as including a separate dummy variable for every student in the regression dataset.³³⁶ As explained in my Initial Report, student fixed effects account for time-invariant factors intrinsic to each student, such as general academic aptitude and time-invariant socioeconomic and demographic factors.³³⁷

191. Dr. Hill asserts that my regressions that include student fixed effects are flawed.³³⁸ More specifically, he claims that those models do not use Emory’s or Dartmouth’s data for identification of the overcharge because these schools provided data only for first years.³³⁹ He also claims that my student fixed effects models primarily only measure price changes for locked-in students.³⁴⁰

192. Dr. Hill’s argument that Emory and Dartmouth do not inform identification of the overcharge in my student fixed effects regressions is unconvincing. Plaintiffs in this case sent discovery requests for complete structured financial aid data for *all* undergraduates.³⁴¹ That my preferred model requires multi-year data per student and Emory and Dartmouth did not produce said

335. See, e.g., WOOLDRIDGE Chapter 14.

336. *Id.*

337. Singer Report ¶245.

338. Hill Report ¶228.

339. *Id.* ¶228.

340. *Id.*

341. See Request No. 7, Plaintiffs’ First Set of Requests for Production of Documents, *Henry v. Brown University*, Case No. 1:22-cv-00125 (N.D. Ill. Sept. 10, 2022).

data does not impugn my modeling choices. My preferred model remains the most accurate representation based on the available evidence. Furthermore, my Defendant fixed effects models, which use Emory's and Dartmouth's data for identification, show that Class Members paid a positive, economically, and statistically significant overcharge.

193. Dr. Hill's assertion that my student fixed effects models primarily only measure price changes for locked-in students is misleading. Dr. Hill bases this argument solely on his claim that my student-fixed effect regressions "primarily rely on price changes for students that are locked in (since the regressions rely entirely on having *at least* two years of data for each student)."³⁴² He does not provide any cite to justify his argument, nor does he explain the logic behind it. Dr. Hill's argument seems to imply (although he never explains it) that, because my student fixed-effects models require that a student appear in the data for more than one year, these models therefore may not be perfectly representative of all Class Members because some Class Members either drop out or transfer and therefore only show up for one year. His argument ignores that singleton students—that is, students who show up for only one year—comprise only 11 percent of the regression data.³⁴³ I note that Dr. Hill's argument should not be misconstrued as suggesting that my student fixed effects models do not consider pricing variation for first-years—for a student to show up as being a second, third, or fourth year, they must have also been a first-year, and therefore do inform the conduct coefficient.

194. In fact, Dr. Hill's critique of my student fixed effects regressions implies that my overcharge estimates are conservative. Dr. Hill claims that, if there is lock-in for first years, this would imply that Effective Institutional Prices for first-years are "more informative about the impact of the alleged conduct (since those students are not locked in) than information in subsequent

342. Hill Report ¶228 (emphasis added).

343. See my workpapers.

years.”³⁴⁴ Based on Dr. Hill’s argument, because my student-fixed effect regressions “primarily rely on price changes for students that are locked in,” and because students that are not locked in would exhibit greater price variation attributable to the conduct, this would mean that my conduct coefficient would in fact be *less* than if all first-years were used for identification.

195. Fundamentally, Dr. Hill’s critiques of my student fixed effects regressions are unavailing because they ignore the importance of controlling for time invariant, student-specific factors that may be unaccounted for by my non-student fixed effects regression models. These “unit” fixed effects enjoy widespread use throughout the statistical literature as a method to avoid potential omitted variable bias when analyzing panel data such as these.³⁴⁵

h. The Presence of Staggered Treatment Underscores the Conservative Nature of My Conduct Coefficient and Resulting Damages

196. Dr. Hill asserts that, because my model uses a single dummy variable to measure the conduct, and because certain schools entered and exited the 568 Group at different times, this condition results in affected students serving as a control for other affected students.³⁴⁶ Dr. Hill argues that this results in scenarios where decreases in Effective Institutional Prices translate into higher damages for 29 percent of school-academic year combinations.³⁴⁷

197. Before responding to Dr. Hill’s critique, I first explain Dr. Hill’s argument. While he casts his critique of my use of a single dummy variable as a “well-understood economic problem,” he offers little detail on that position. In fact, he does not provide any citation to support his argument. I interpret Dr. Hill’s criticism to mean that varying entry and exit periods for various Defendants

344. Hill Report ¶228.

345. Muhammad Al Amin, *Panel Data Using Stata: Fixed Effects and Random Effects*, PRINCETON UNIVERSITY LIBRARY RESEARCH GUIDES, <https://libguides.princeton.edu/stata-panel-fe-re> (“Entity fixed effects account for unobserved heterogeneity across entities (e.g., individuals, firms, countries) that is constant over time but varies between entities. This is the most frequently used model in panel data analysis.”). See also WOOLDRIDGE Chapter 14.

346. Hill Report ¶233.

347. *Id.*

(e.g., Yale) creates a staggered treatment issue. Without accounting for this issue, some affected members in one period can serve as a control for affected members in another period, instead of only unaffected members serving as controls for affected members.

198. Staggered treatment conditions lie at the forefront of much of the difference-in-difference literature. On this point, Dr. Hill's claim that staggered treatment is a "well-understood economic problem" stands in conflict with the literature. On the contrary, researchers acknowledge the difficulty staggered treatment can pose under certain conditions.³⁴⁸ Dr. Hill offers little insight to the problem, other than referencing the Callaway and Sant'Anna paper on this topic.

199. The Callaway and Sant'Anna paper is well known in the econometric literature. I take no issue with Dr. Hill's reference to this paper and take his argument into consideration. Yet the authors explain: "We concentrate our attention on DiD with staggered adoption, i.e., to DiD setups such that once units are treated, they remain treated in the following periods."³⁴⁹ *First*, as I explained in my Initial Report, I noted that I used the before-and-after approach, not the difference-in-differences method. *Second*, the setup that Callaway and Sant'Anna contemplate differs from the instant case. Defendants entered and exited the conspiracy and could enter again, as Yale did.

348. Andrew Goodman-Bacon, *Difference-In-Differences with Variation in Treatment Timing*, 225(2) JOURNAL OF ECONOMETRICS 254–277, 255 (2021) [hereafter Goodman-Bacon (2021)] ("In contrast to our substantial understanding of canonical 2x2 DD, we know relatively little about the two-way fixed effects DD when treatment timing varies. We do not know precisely how it compares mean outcomes across groups. We typically rely on general descriptions of the identifying assumption like 'interventions must be as good as random, conditional on time and group fixed effects' (Bertrand et al., 2004, p. 250). We have limited understanding of the treatment effect parameter that regression DD identifies. Finally, we often cannot evaluate how and why alternative specifications change estimates."). *See also* Seth M. Freedman, Alex Hollingsworth, Kosali I. Simon, Coady Wing, and Madeline Yozwiak, *Designing Difference in Difference Studies with Staggered Treatment Adoption: Key Concepts and Practical Guidelines*, NATIONAL BUREAU OF ECONOMIC RESEARCH, NBER Working Paper No. 31842 (November 2023) ("Difference-in-difference (DID) estimators are a valuable method for identifying causal effects in the public health researcher's toolkit. A growing methods literature points out potential problems with DID estimators when treatment is staggered in adoption and varies with time. Despite this, no practical guide exists for addressing these new critiques in public health research.").

349. Brantly Callaway and Pedro H.C. Sant'Anna, *Difference-in-Differences with Multiple Time Periods*, 225(2) JOURNAL OF ECONOMETRICS 200–30, 201 (2021).

200. Elsewhere in his report, when offering original empiricism to support an affirmative opinion, Dr. Hill states:

Defendants entered and exited the 568 Group at various times, and some schools entered or exited multiple times. This variation in entry and exit dates poses a problem for a standard difference-in-differences model, which is predicated upon a single event that occurs at a single point in time. To handle this econometric problem, I use the staggered difference-in-differences model developed by Callaway and Sant'Anna (2021) rather than a difference-in-differences model with a single dummy variable for the alleged conduct. This approach has some limitations, but with staggered entries and exits, it is important to use it because the standard difference-in-differences approach is flawed.³⁵⁰

Yet when attempting his “correction,” Dr. Hill limits his analysis to only seven defendants. I reject such selective and unsupported *ad hoc* criteria as Dr. Hill imposes. I see no reason to artificially limit the scope of the conspiracy to only seven defendants, and Dr. Hill offers none other than that doing so allows him to apply his preferred technique. But Dr. Hill has it backwards: the analysis should apply to the issue at hand, not vice versa.

201. Goodman-Bacon explains the staggered treatment issue within the context of DiD model succinctly in his 2021 paper:

Some use units treated at a particular time as the treatment group and untreated units as the control group. Some compare units treated at two different times, using the later-treated group as a control before its treatment begins and then the earlier-treated group as a control after its treatment begins. The weights on the 2x2 DDs are proportional to timing group sizes and the variance of the treatment dummy in each pair, which is highest for units treated in the middle of the panel...When treatment effects do not change over time, TWFEED [two-way fixed effects difference-in-differences] yields a variance-weighted average of cross-group treatment effects and all weights are positive.³⁵¹

From an economic viewpoint, given the nature of the alleged conspiracy, one expects stable treatment effects from the Overarching Agreement, given that members joined the group to follow its stated goals. Despite Defendants’ Experts attempts to theorize various sources of treatment heterogeneity, the scope and purpose of the 568 Group remained constant over time. Whether the group applied

350. Hill Report ¶89, citations omitted.

351. Goodman-Bacon (2021).

various elements of the alleged conspiracy with varying rigor, just as one applies various condiments in preparing a dish, is irrelevant because the end result remained the same.³⁵²

202. To shed additional light on Dr. Hill's misplaced criticism, consider the following example. Suppose various individuals were given a vaccine against a certain virus. Suppose further that some individuals had already contracted the virus previously and had recovered. As a result, these individuals had developed antibodies against the virus. Alternatively, suppose some individuals had received previous treatments of the virus years ago, but required a new booster.

203. Ideally, in such a setting, we would conduct a randomized controlled trial. Given the observational nature of economic data, however, we cannot. Suppose we perform a before and after analysis, similar to what I have applied herein. Including individuals who had antibodies from a previous infection as a control would dilute any causal effects of the vaccine, because we are not comparing a person with vaccine-induced antibodies with a person without antibodies at all. As such, we would *underestimate* the efficacy of the vaccine, rendering our causal effect conservative. The same logic applies here. To the extent that previous membership in the conspiracy reduced a Defendant's financial aid generosity subsequent to its departure from the 568 Group (just as antibodies from a previous infection or a previous vaccine might continue to offer protection after the expected cessation of said protection), treating such a Defendant as a "clean" benchmark would understate the effect of the conspiracy.

352. For a further elaboration on these issues in DiD models, see Clement de Chaisemartin and Xavier D'Haultfoeuille, *Two-way fixed effects estimators with heterogeneous treatment effects*, 110(9) AMERICAN ECONOMIC REVIEW 2964–2996, 2965 (2020) (“[T]he ‘control group’ in some of those comparisons may be treated at both periods. Then, its treatment effect at the second period gets differenced out by the DID, hence the negative weights. The negative weights are an issue when the ATEs are heterogeneous across groups or periods.”). Goodman-Bacon (2021) explains negative weights as follows: “Negative weights only arise when average treatment effects vary over time. The DD decomposition shows why: when already-treated units act as controls, changes in their outcomes are subtracted and these changes may include time-varying treatment effects. *This does not imply a failure of the design in the sense of non-parallel trends in counterfactual outcomes*, but it does suggest caution when using TWFE estimators to summarize treatment effects.” (emphasis added).

i. My Results Are Not Subject to Selection Bias, Contrary to Dr. Hill's Unsupported Claims

204. Dr. Hill claims that limiting the regression sample to only students who (1) received positive institutional grant aid and (2) did not get a full ride introduces selection bias into my regression.³⁵³ In attempting to support his argument, Dr. Hill does not rely on any actual data. Instead, he creates two “toy” problems, or simplified examples used to illustrate a methodology. Specifically, Dr. Hill uses a hypothetical example of a student who switches in one year from paying a positive Effective Institutional Price to then getting a free ride in their second year, with the second year being subject to the Challenged Conduct. Dr. Hill averages this student’s Effective Institutional Price with another student who paid \$30,000 in both years, and he finds that such a scenario would show that Effective Institutional Prices increased as a result of the Challenged Conduct.

205. Based on his hypothetical example, for which he provides no link to the actual data, he claims that, by limiting the regression sample to only students who received at or less than 95 percent of the cost of attendance, I purportedly introduce a selection bias similar to said scenario.³⁵⁴ In other words, he creates a hypothetical scenario then, without verifying it applies to the actual data in any way, he speculates that it does apply. Because his argument does not reveal any relationship between this manufactured toy problems and actual data, his position has no application to this case. Moreover, because he could not muster a single such example from the actual world, his critique for this reason is untethered to my analysis. Further, Plaintiffs’ counsel assigned me the task of measuring harm to the Class as defined. Had I included non-Class Members in my analysis, I expect Defendants’ Experts would have levied various criticisms against doing so.

353. Hill Report ¶234, ¶236.

354. *Id.* ¶234, ¶237.

j. The Presence of Elements of the Challenged Conduct in Both Benchmark and Conspiracy Periods Indicate that My Coefficient Underestimates the True Classwide Harm from the Conduct

206. Dr. Stiroh claims that my regression cannot distinguish between harm caused by an individual element of the six components of the Challenged Conduct.³⁵⁵ According to Dr. Stiroh, if any less than all six components were found to have violated the law, she claims that I would be unable to disentangle the effect of only those unlawful conducts from the effects of other Challenged Conduct but not unlawful conducts on Effective Institutional Prices. Dr. Stiroh also argues that other elements, such as sharing COFHE data³⁵⁶ and utilizing need-based aid as the primary aid form,³⁵⁷ existed both during and outside of the conduct period. She concludes that the presence of the same conduct in both conspiracy and benchmark period inhibits the recovery of any causal effect from the Challenged Conduct.

207. This critique is misguided for at least three reasons. *First*, I understand from counsel that the elements of the Challenged Conduct comprise an alleged single overarching conspiracy. As such, Dr. Stiroh's attempt to separate elements of the same claim is akin to attempting to attribute a cheetah's speed to one of its four legs. Just as the cat's velocity depends on the concerted contribution of each of its four legs, the combination of various elements of the Challenged Conduct effectuated the conspiracy.

208. *Second*, the attempt to separate out these elements causes Dr. Stiroh to commit a logical fallacy, upon which she bases her claim that the presence of elements of the Challenged Conduct both before and during the Class Period obviates any causal relationship between an increase in price and the conduct at issue. Each element of an orchestra can play individually, but only when

355. Stiroh Report ¶188.

356. *Id.* ¶186.

357. *Id.* ¶185.

combined together do they create a symphony. Dr. Stiroh's faulty logic would argue that the musical instruments could not have been the cause of the symphony because they existed both before and after the orchestra began playing in tandem and as a whole.³⁵⁸

209. *Third*, citing various other literature, Boswijk et al. observe that elements and effects of a cartel can exist beyond the formal dates that bound a cartel's formalistic existence:

The formal cartel dates need not coincide with the point(s) in time at which the anticompetitive effects produced by the cartel violation began or ended. The collusive agreements may take a while to take effect, become ineffective before its members ultimately disband, or rather have lasting effects—including unilateral incentives for former cartellists to keep post-cartel prices up in order to mask their conspiracy (see Harrington, 2004). Also, between periods of coordinated high prices, the cartel spell can include temporary price wars or reversionary episodes due to internal tensions, from which the cartel regathered.³⁵⁹

As such, I reject Dr. Stiroh's criticism as misunderstanding both the elements of the alleged conspiracy and the nature and duration of their effects.

210. Neither Dr. Hill nor Dr. Stiroh observe that the presence of any elements of the conspiracy in the benchmark periods would only serve to *dilute* the effects of the alleged conspiracy, resulting in a lower coefficient. That the coefficient remains statistically and economically significant even in the possible presence of conduct contamination of the benchmark periods is a testament to the effect of the alleged conspiracy. In that way, my results then offer a conservative estimate of the alleged conspiracy's effects.

358. As observed in the *State of NY vs. Facebook* Amicus Brief that I co-signed with other antitrust economists, including Nobel Laureate Joseph Stiglitz, an orchestral concert provides an appropriate analogy in such a case. "[E]ach instrument on its own produces a sound, but a symphony represents far more than the sum total of separate instruments. Because the sounds of the instruments reinforce each other to produce the symphony, they cannot be separated into individual sonatas and still produce the same result." *State of NY v. Facebook*, Case No. 21-7078, Brief of Economists as Amici Curiae in Support of Plaintiff-Appellants and Reversal (D.C. Cir. Jan. 28, 2022) <https://www.cohenmilstein.com/wp-content/uploads/2023/07/New-York-v-Facebook-Economists-Amicus-Filed-01282022.pdf>.

359. H. Peter Boswijk, Maurice J. G. Bun, and Maarten Pieter Schinkel, *Cartel Dating*, 34 JOURNAL OF APPLIED ECONOMETRICS 26-42, 26 (2017).

B. Classwide Evidence Demonstrates that the Overcharge Would Have Impacted All or Nearly All Class Members

211. In Part III.B of my Initial Report, I presented the second step of my two-step methodology for demonstrating class wide impact. The first step involved showing that the Challenged Conduct resulted in a generalized overcharge to Effective Institutional Prices paid by Class Members through both qualitative and quantitative evidence, including my overcharge regression model. As described in the previous section, none of Defendants' Experts disprove the reliability of my overcharge model in showing a generalized overcharge to Class Members. The second step of my two-step methodology, described in more detail below, involves demonstrating that the *generalized* overcharge caused by the Challenged Conduct resulted in harm via higher Effective Institutional Prices *to all or nearly all* Class Members.

212. Dr. Stiroh is Defendants' only expert to offer an opinion regarding my second step of demonstrating classwide impact. Dr. Stiroh highlights various elements of the Challenged Conduct that she asserts would have prevented common impact among Class Members. She then critiques the three analyses that I provide as quantitative evidence for classwide impact—in-sample prediction, a common-shock analysis, and price structure regressions. Below, I explain the flaws in Dr. Stiroh's arguments.

213. It is worth noting that in Part III.B.2.b of my Initial Report, I provided qualitative evidence showing that Defendants' goals of maintaining horizontal and vertical equity are consistent with the existence of an Effective Institutional Price structure. These equity considerations mean that "families of similar socioeconomic status and size should receive similar aid."³⁶⁰ Defendants' Experts do not dispute this qualitative evidence that I brought forward.

360. Singer Report ¶269.

1. Dr. Stiroh's Principle-by-Principle Approach to Assessing Classwide Impact Is Fundamentally Flawed

214. Dr. Stiroh claims that there is no economically plausible mechanism through which Defendants' alleged agreement on core principles would result in classwide impact.³⁶¹ As support for her claim, she goes through each component of the Challenged Conduct one-by-one and assesses whether, in a world absent the individual component, all or nearly all Class Members would have paid lower Effective Institutional Prices.³⁶²

215. Dr. Stiroh's approach is misguided at a high level because my assignment was not to assess whether each *separate* component of the Challenged Conduct would have harmed all or nearly all Class Members; rather, I analyzed whether the Challenged Conduct *as a whole* would have resulted in classwide impact. I do precisely that, by showing that a generalized overcharge in Effective Institutional Prices resulting from the Challenged Conduct would have impacted all or nearly all Class Members via qualitative evidence and economic and econometric analyses, including in-sample prediction and a price structure analysis.

216. Dr. Stiroh's separate assessment of each principle of the Challenged Conduct also contains flawed arguments. Dr. Stiroh argues that additional professional judgement would not necessarily have benefited Class Members, because there is "no class-wide method to establish in which additional or different cases schools would have opted to exercise professional judgement that they did not exercise in the actual world[.]"³⁶³ This point is irrelevant. Absent the alleged conspiracy, Defendants would have used the same method, including professional judgement, to set financial aid

361. Stiroh Report §V.

362. *Id.* ¶81.

363. *Id.* ¶¶96-97.

as they did in the years in which they did not participate in the Challenged Conduct.³⁶⁴ Whatever methods Defendants used outside of the Challenged Conduct, whether an increase in professional judgment or an idiosyncratic modification of the IM, resulted in lower Effective Institutional Prices for all or nearly all Class Members when compared to the Class Period. Dr. Stiroh’s discussion of professional judgment obfuscates this point. Further, the qualitative evidence suggests that professional judgment was not a predominant factor in determining EFCs—it was treated as “the exception, rather than the rule.”³⁶⁵ Dr. Long agrees with this point, stating that “[o]ver time, however, the goal is for frequently recurring circumstances to be accounted for by the baseline rules or methods, reserving professional judgment adjustments for one-off or unique situations.”³⁶⁶ Accordingly, if there were some discretion in professional judgment, this discretion would not be expected to alter the classwide impact of a generalized overcharge.

217. Dr. Stiroh claims that merit aid is not used by all Defendants, even when they are outside of the 568 Group, and that by taking primacy over need-based aid, increasing merit aid could decrease need-based aid.³⁶⁷ She further argues that there would be no way to tell which students would benefit from increased merit aid.³⁶⁸ These arguments are without merit because they assume a fixed pot of aid funding, which has no basis, and because they are not pertinent to classwide impact. As I previously stated, I did not opine that the primacy of need-based aid was driving the changes in

364. Even if Defendants were to choose a “base” price in each academic year, and then vary this price by student via subsequent individualized negotiations, this would still result in anticompetitive effects. This is because the artificial inflation in base price resulting from the Challenged Conduct would shift all subsequent individualized prices. *See, e.g.,* AMERICAN BAR ASSOCIATION, *ECONOMETRICS: LEGAL, PRACTICAL, AND TECHNICAL ISSUES* (ABA BOOK PUBLISHING 2nd ed. 2014) [hereafter *ABA ECONOMETRICS*] at n. 18 (“Class members do not have to buy the same homogeneous product at a uniform price to be found to be in the same class. Classes have been certified that include different members that purchased different products or paid substantially different prices for the products or services they purchased. However, this generally requires statistical evidence showing that there is a stable structural relationship between the prices paid by the different class members.”).

365. GTWNU_0000347907 at -914.

366. Long Report ¶124.

367. Stiroh Report ¶¶92-93.

368. *Id.*

the Effective Institutional Price. My regression shows that the Effective Institutional Price was lower for all or nearly all Class Members when Defendants were not engaged versus when they were engaged in the Challenged Conduct. This finding would not logically or economically require that Defendants increased merit aid in the but-for world. If Defendants used merit aid both during the Class Period and during non-participatory periods, then whether they, or all of them, would adopt or use merit aid more intensively is not pertinent. In a more competitive but-for world, some Defendants may have increased merit aid, some may have increased need-based aid, and some may have increased both. Ultimately, any of these outcomes would produce greater institutional gift aid and lower prices. I understand that Defendants had the ability to devote additional funds to financial and/or merit aid. Thus, if some Defendants would have increased merit aid in the but-for world, they would have had enough resources to give more need-based aid as well.³⁶⁹

218. Dr. Stiroh argues that the effect of the use of COFHE data is impossible to assess on a classwide level, because it is aggregated and averaged data.³⁷⁰ She also claims that the use of such data would not make all Class Members worse off.³⁷¹ It is important to note that common impact does not hinge on the use of COFHE data alone. It is not sine qua non, nor did I ever make this claim in my Initial Report. Economic theory makes clear that sharing competitively sensitive data and information as was part of the Challenged Conduct can both cause higher prices in and of itself and also facilitate the enforcement of a price-fixing cartel.³⁷² Furthermore, Dr. Stiroh ignores the degree

369. Bulman Report §V.C.

370. Stiroh Report ¶103.

371. *Id.* ¶¶104-105.

372. See, e.g., Joseph E. Harrington, *Detecting Cartels* (2005) at 2 (“The means of coordination is some form of direct communication and, indeed, many cartels have been detected by virtue of evidence of that communication.”); Margaret C. Levenstein & Valerie Y. Suslow, *What Determines Cartel Success?*, 44(1) JOURNAL OF ECONOMIC LITERATURE 43-95, 44 (2005) (“Cartels much prefer to develop the means to monitor each other's behavior in order to deter or physically prevent cheating, rather than resorting to expensive punishments such as price wars. Designing effective monitoring mechanisms takes place over time as cartels learn about both their competitors and their customers, and then refine the organizational structure to provide the necessary incentives and information to sustain cooperation.”).

of aggregation. For instance, the Bluebook breaks down details about “Parent Contribution” and “Net Price” broken down by groupings of family income.³⁷³ “The Financial Aid Supplement, similarly, extends the reach of unit-record data to enable comparison of average grants by income group across groups of member institutions.”³⁷⁴ While the COFHE data may not have broken down contributions and price by student, that it is disaggregated by income level is still relevant for purposes of classwide impact, specifically since incomes are one of the primary factors that Defendants assessed for purposes of determining EFCs.

219. In my Initial Report, I explained that vertical and horizontal equity would be consistent with classwide impact. Dr. Long acknowledges the importance of horizontal equity, in which students in similar financial circumstances receive similar amounts of aid, as a goal of Defendants and “many other schools.”³⁷⁵ Dr. Long provides an example where “two students whose families report the same annual income and net worth, but one family lives in Massachusetts (which has the fourth-highest cost of living for all states) while the other lives in Mississippi (which has the fifth-lowest cost of living for all states). Despite having the same income, these two families may have different financial circumstances driven by differences in the cost of living they experience.”³⁷⁶ Dr. Long goes onto state that universities providing need-based aid develop methodologies and guidelines to ensure that “those with the greatest level of financial need receive the most financial aid (i.e., vertical equity), and students with similar levels of financial need receive similar amounts of financial aid (i.e., horizontal

373. See Columbia_00277826 (a 55-page document titled “Freshman Financial Aid/Admissions Survey – Class of 2017 – Entering 2013 – (Bluebook XXIX) – Consortium on Financing Higher Education – February 2014”). Pages 5-6 state that “*Bluebook* tracks a wide variety of statistics related to financial aid and admissions” that concern, *inter alia*, matriculants receiving aid, the net price faced by aided students, parent contributions, average net prices, and family incomes. Pages 9-10 state that “This [Bluebook] includes data provided by the financial aid and admissions offices of the 31 COFHE members and five additional institutions that have traditionally participated in this study. . . Using unit data, COFHE calculated the counts, means, and other statistics usually reported in *Bluebook*. These figures were then returned to each institution – along with the logic used to calculate them – for review and sign-off.”

374. CORNELL_LIT0000038218.

375. Long Report ¶85.

376. *Id.* ¶119.

equity).”³⁷⁷ Defendants had used and would have continued to use a formulaic need-based approach to awarding aid absent the Challenged Conduct, suggesting that an increased level of aid would have been transmitted to all students in the but-for world, regardless of income level.³⁷⁸

220. Finally, Dr. Stiroh claims that common impact cannot be established because using the Base IM would not result in uniformly lower EFCs.³⁷⁹ This is a strawman argument. There is no need to propose an alternative methodology. The benchmark does not contemplate the use of the Base IM, but rather whatever modification of the IM or FM Defendants would have used absent an alleged two-decade conspiracy to share information and collude on principles and a formula. The evidence and analyses in my reports have shown that unfettered competition would have prompted the development of a more generous “alternative” methodology or set of more generous methodologies in the but-for world. The purpose of my regression analysis was not to show uniformity in the amount of lowered Effective Institutional Prices, but rather that every Class Member would have benefited, to some degree, from the absence of the Challenged Conduct. This conforms with the expectation that Defendants would be risk averse in the but-for world. That is, with Defendants competing with each other for students and subsequently more aid being distributed, it would be rational, risk-minimizing behavior to increase the aid to all or nearly all students. As Dr. Stiroh highlights when discussing universities benchmarking their own practices against competitors, in a world with unfettered competition, one would expect Defendants to compete on net price and to offer more generous institutional grant aid.³⁸⁰

377. Long Report ¶118.

378. Dr. Long opines that both Defendants and non-Defendants implemented need analysis systems, implying that Defendants would have continued to use need analysis systems in a more competitive world. *Id.*

379. Stiroh Report ¶108, ¶112.

380. *Id.* ¶102.

221. Dr. Stiroh also raises a separate issue by claiming that “[w]hen a challenged agreement does not cause alleged conspirators to take actions in the same direction (i.e., reduce output or raise price), there is no economic basis to expect that the agreement had common impact on a proposed class.”³⁸¹ This argument is unfounded for two reasons.

222. *First*, it contradicts both the qualitative and quantitative evidence I have marshalled showing that each Defendant had artificially inflated their Effective Institutional Prices as a result of the Challenged Conduct. In my Initial Report, I demonstrated that each Defendant had artificially inflated its Effective Institutional Prices during years in which they participated in the Challenged Conduct through my Defendant-specific in-sample prediction analysis.³⁸²

223. *Second*, it is not true as an economic matter. There are many different dimensions to the Challenged Conduct that do not necessarily equate to Defendants raising prices or decreasing output perfectly in parallel.³⁸³ For instance, the qualitative and quantitative evidence suggest that consensus on the CM resulted in a price floor for EFCs, and that the resulting artificial inflation in EFCs from implementation of this floor artificially inflated Effective Institutional Prices. EFCs could still vary by Defendant under this consensus on CM, as explained in Part II.A above, and so could Effective Institutional Prices. Even with variation in Effective Institutional Prices across Defendants, said prices would still be subject to the artificial overcharge resulting from the EFC floor.

381. Stiroh Report ¶71.

382. Singer Report Table 13.

383. *Price Fixing, Bid Rigging, and Market Allocation Schemes: What They are and What to Look For*, U.S. DEPARTMENT OF JUSTICE (revised Feb. 2021), <https://www.justice.gov/d9/pages/attachments/2016/01/05/211578.pdf> at 2 (“It is not necessary that the competitors agree to charge exactly the same price, or that every competitor in a given industry join the conspiracy.”).

2. Econometric Evidence Confirms That the Generalized Effective Institutional Price Overcharges Were Paid by All or Nearly All Class Members

224. In my Initial Report, I empirically demonstrated that the artificial overcharge in Effective Institutional Prices resulting from the Challenged Conduct can be shown to have impacted all or nearly all Class Members during the Class Period. I do this using two sets of analyses. *First*, I presented an in-sample prediction analysis, which showed that 97 percent of Class Members are predicted to have been impacted via paying artificially inflated Effective Institutional Prices than they would have absent the Challenged Conduct.³⁸⁴ *Second*, I presented price structure analyses, which showed that a Class Member's Effective Institutional Prices are highly correlated with the Effective Institutional Prices paid by other Class Members.³⁸⁵ The latter analyses show that a generalized effect across a set of schools or within a school, as my regression modelling has shown, would likely be experienced across students within each school given that their prices are correlated with each other, and so if the average moves down, I would expect nearly all prices to move down with the average. Below, I highlight Dr. Stiroh's critiques of both sets of analyses, and I explain the flaws in her reasoning.

a. Dr. Stiroh's Critiques of my In-Sample Prediction Methodology are Without Merit

225. In my Initial Report, I used in-sample prediction to demonstrate that all or nearly all Class members were negatively impacted by the Challenged Conduct.³⁸⁶ My methodology is rooted

384. Singer Report §III.B.1.

385. *Id.* §III.B.2.a.

386. Singer Report §III.B.1.

in basic econometrics; to dispute the validity of its application, as Dr. Stiroh does, is contrary to standard economic principles.³⁸⁷

226. Dr. Stiroh claims that my in-sample prediction “is fundamentally flawed and econometrically invalid.”³⁸⁸ Her argument is erroneous because (1) it ignores the reality of the data and the first step of the in-sample prediction method, and (2) it is predicated on presenting a basic statistical characteristic of regression predictions as a flaw: that approximately 50 percent of observations in a regression model will be above the regression (or in-sample prediction) line, and 50 percent of observations will be below the regression line. As I demonstrate below, the in-sample prediction method is functioning correctly and provides reliable support of classwide impact.

227. My in-sample prediction analysis is based on a two-step process. First, I calculate the generalized overcharge of Effective Institutional Prices resulting from the Challenged Conduct. Second, I use in-sample prediction to measure how the generalized overcharge impacts each Class Member’s Effective Institutional Prices. This approach to in-sample prediction is based upon the Federal Judicial Center’s *Reference Manual on Multiple Regression*,³⁸⁹ and it has been cited in several

387. See WOOLDRIDGE at 190 (“Sometimes, it is useful to examine individual observations to see whether the actual value of the dependent variable is above or below the predicted value; that is, to examine the residuals for the individual observations. . . . Residual analysis also plays a role in legal decisions.”). See also Daniel L. Rubinfeld, *Reference Guide on Multiple Regression*, 3 REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 303–357, 308 (2011) [hereafter REFERENCE MANUAL].

388. Stiroh Report ¶197.

389. REFERENCE MANUAL at 308.

recent antitrust cases wherein economists have used this same in-sample prediction to demonstrate Common Impact.³⁹⁰ The in-sample prediction method can be summarized as follows:³⁹¹

Step 1: Specify a regression model to estimate *general* overpayments across Class members. The coefficient on the variable of interest represents the *general* or aggregate overpayment to Defendants from Class Members during the Class Period. If the general overpayment is economically and statistically significant, proceed to step two.

Step 2: Compute the “but-for” price for each observation in the data. This is done by using the overcharge regression to calculate estimated Effective Institutional Prices associated with each Defendant-Class Member-academic year *but-for* the Challenged Conduct. This is done by using standard regression prediction methods, after manually setting the effect of the variable of interest (i.e., the conduct variable) to zero. This simulates the price but-for the Challenged Conduct to each Class Member, based on a common econometric model.

Step 3: Compare the actual price paid to the “but-for” price. The difference between the actual and “but-for” price is the overpayment estimate for each Class Member-academic year.

228. Dr. Stiroh’s criticisms are fatally flawed because she ignores Step 1. If it were the case that Effective Institutional Prices were not artificially suppressed by the conduct, then the analysis would end there. There would be no aggregate overpayment to assess on a per-student basis. In that case, the in-sample method would not be advisable, and thus should not be used, as all it will do is demonstrate statistical noise.

229. Instead, contrary to standard practice, Dr. Stiroh begins at Step 2 with an assumed zero general artificial overcharge.³⁹² Under this scenario, she claims that my in-sample prediction would be expected to show 93.75 percent of Class Members impacted based on random chance

390. *In Re Broiler Chicken Growing Antitrust Litigation (No. II)*, 6:20-MD-02977-RJS-CMR (E.D. Ok May 8, 2024) (Memorandum Decision and Order Granting Plaintiffs’ Motion for Class Certification and Denying Defendant’s Motion to Exclude) [hereafter *In Re Broiler Chicken Grower Class Certification*] at 40 (“The in-sample prediction method is a standard technique used to test whether the impact of an antitrust conspiracy is widespread.”); *id.* at n. 242 (citing *Olean Wholesale Grocery Coop., Inc. v. Bumble Bee Foods, LLC*, 31 F.4th 651, 672 (9th Cir. 2022); *In re Capacitors Antitrust Litig. (No. III)*, Case No. 17-md-02801, 2018 WL 5980139, at *7–9 (N.D. Cal. Nov. 14, 2018); *In re Domestic Drywall Antitrust Litig.*, 322 F.R.D. 188, 217 (E.D. Pa. 2017); *In re Korean Ramen Antitrust Litig.*, Case No. 13-cv-04115, 2017 WL 235052, at *6 (N.D. Cal. Jan. 19, 2017); *In re Broiler Chicken*, Case No. 16-cv-8637, 2022 WL 1720468, at *10, *13 (N.D. Ill. May 27, 2022).

391. *See* n. 389, *supra*.

392. *Id.* ¶198.

alone.³⁹³ Her finding is analogous to flipping a coin four times: the probability that at least one flip will turn up heads is $1 - (0.5)^4 = 0.9375$. This argument is predicated on the false premise that the generalized overcharge equals zero. If the generalized overcharge were zero, the analysis would stop there.

230. Rather, both qualitative and quantitative evidence indicate that the Challenged Conduct resulted in a generalized overcharge to Effective Institutional Prices. Therefore, for every year that a Class Member was subject to the *generalized* overcharge, the higher the likelihood that said Class Member would in fact suffer an overcharge via a higher Effective Institutional Price in any year during the Class Period, consistent with my findings.

231. My primary Effective Institutional Price regression model explains approximately 87 percent of the variation in Effective Institutional Price, over and above that which could be explained by a naïve estimate based on the mean Effective Institutional Price alone. An R-squared of 87 means that only 13 percent of price variation is unexplained. Indeed, in a well-functioning regression model, we would *expect* the unexplained variation to fall on either side of the estimated regression line evenly. That is, the 50-50 split is the benchmark. The fact that 96 percent of observations in the in-sample prediction method I use show a but-for Effective Institutional Price *below* the actual Effective Institutional Price (that is, demonstrate impact), when the *expected value* in the case of no overcharge is 50 percent, demonstrates that the model is detecting impact beyond random chance. It is finding a signal and not noise.³⁹⁴

393. *Id.*

394. In my reply workpapers, I provide revised versions of my in-sample prediction showing the percent of Class Members impacted at least once, and the in-sample prediction overcharge calculation by Defendant, both of which were provided in my Initial Report. I use the revised regression data that is adjusted to incorporate all of Dr. Hill's data-processing adjustments marked as "Accepted" = "Y" outlined in Table 5. I find that 96 percent of Class Members paid an artificially inflated Effective Institutional Price at least once. This is approximately equal to the 97 percent in-sample prediction result that I provided in Table 12 of my Initial Report. I also continue to find that each Defendant had artificially inflated Effective Institutional Prices using the Defendant-specific in-sample prediction methodology presented in Table 13 of my Initial Report.

232. Furthermore, my in-sample prediction analysis conservatively assumes that the unexplained variation may allow Class Members to escape overcharges when, in reality, there is no reason to conclude that variation would have been different in the but-for world. In this sense, the model underestimates impact: If my regression could predict with perfect accuracy the Effective Institutional Price every Class Member would pay during every academic year, then any overcharge would show 100 percent impact. By attributing the unexplained error to the ability of a Class Member to escape injury, the model assesses if there exists a large number or segment of students that consistently paid lower Effective Institutional Prices than what the regression model would predict. In this case, the model shows that virtually all Class Members paid artificially inflated Effective Institutional Prices.

233. In addition, my in-sample analysis assumes that any student who actually paid less than the predicted but-for Effective Institutional Price did so because they benefitted from the Challenged Conduct. In reality, the reason why they show as having paid less than the predicted but-for Effective Institutional Price is for the reason explained above—my model does not perfectly explain every factor influencing Effective Institutional Prices. There is no basis for assuming other factors not accounted for by my model would differ in the actual versus the but-for world. In other words, any student who shows up as having paid less than the predicted but-for Effective Institutional Price would likely have paid *even less* without the Challenged Conduct. All of the factors that Dr. Stiroh points out as having a differential impact on students—professional judgment, packaging special circumstances, etc.—would all be expected to have occurred in the same way in both in the actual and the but-for world.

234. I highlight the flaw in Dr. Stiroh’s argument by demonstrating that Class Members were impacted on a “harmed on net” standard. To be impacted “on net,” a Class Member would have

had to have been overcharged when summing across all of his or her academic years during the Class Period. That is, if I sum but-for total Effective Institutional Prices over all transactions for each Class Member and compare it to the sum of their actual Effective Institutional Prices, net harm occurs when the Class Member would have paid less *on net* (across all academic years during the Class Period) absent the Challenged Conduct. Finding impact under this methodology requires a more general overpayment, and thus it is less likely that its results would be due to statistical chance, as Dr. Stiroh asserts. Put another way, Dr. Stiroh claims that my in-sample prediction methodology could be flawed due to a scenario where a student has a fifty-fifty chance of being impacted by paying a higher Effective Institutional Price. If this were the case, then any predicted overcharges per student would be expected to be cancelled out by predicted undercharges, since actual values would show up equally on both sides of the regression line, and this would render the sum of a student's predicted overcharges across all academic years in the Class Period to zero. If one finds a high proportion of students suffering from positive overcharges across all academic years subject to the Challenged Conduct, such a result is less likely the result of statistical noise.

235. Using this “harmed on net” methodology, I find that 95 percent of Class Members overpaid via artificially inflated Effective Institutional Prices across all academic years they were subject to the Challenged Conduct. This result runs counter to Dr. Stiroh’s assertion that my results are “mechanical.”³⁹⁵

236. Dr. Stiroh claims that because I had improperly assigned the academic years for Duke, Northwestern, and Vanderbilt in my Initial Report, prior to Dr. Hill’s adjustment to these Defendants’ data, my in-sample prediction improperly finds higher rates of impact for these Defendants than the

395. *Id.* ¶201.

average Class Period observation.³⁹⁶ This argument is without merit. My in-sample prediction methodology is not flawed because of a mistiming in these Defendants' academic years. As already stated, in-sample prediction is predicated on the existence of a generalized overcharge. That I predict impact in years where these three Defendants had not participated is due to the underlying misassignment of their academic years, not due to a flaw in my methodology.

237. Dr. Stiroh goes on to apply my in-sample prediction to the six Defendants that her flawed Defendant-specific regressions show as having either a negative coefficient, or else a positive, but not statistically significant, coefficient.³⁹⁷ She finds that *some* of these Defendants report a high proportion of Class Members impacted based on her model. Dr. Stiroh asserts that this highlights the “fundamental flaw” in my methodology.³⁹⁸ This analysis does not highlight anything other than the errors in Dr. Stiroh's own logic. It assumes that Dr. Stiroh's Defendant-specific regression models are valid. As I described in Part III.A.2.a, Dr. Stiroh's Defendant-specific regressions are flawed in that they ignore the limitation of the data coverage relative to the Class Period, and because they alter my identification approach. Hence, Dr. Stiroh's flawed Defendant-specific regression produce erroneous in-sample prediction results.

238. In fact, this analysis highlights the flaws in Dr. Stiroh's own logic. Three of the six Defendants she analyzes—Chicago, Notre Dame, and Penn—all show impact for over 90 percent of Class Members (at least once). These three Defendants also happen to be the three Defendants that have a positive (albeit not statistically significant) overcharge in Dr. Stiroh's Defendant-specific regressions. For the other Defendants for which Dr. Stiroh obtains negative conduct coefficients using

396. *Id.* ¶201. Dr. Stiroh states: “For example, Dr. Singer flags 71.8 percent of student-academic year observations at Duke as ‘impacted,’ 68.1 percent at Northwestern, and 70.6 percent at Vanderbilt in academic-years before the Class Periods started.”).

397. *Id.* ¶202.

398. *Id.*

her flawed Defendant-specific regressions—Caltech, Columbia, and Vanderbilt—her in-sample analysis finds significantly lower Class Member impact percentages (22 percent to 79 percent of Class Members impacted at least once). These results are consistent with the validity of my methodology in measuring Class Member impact—students subject to the Challenged Conduct are highly likely to have been impacted; students not subject to the Challenged Conduct are not likely to have been impacted.

b. Dr. Stiroh’s Critiques of My Price Structure Regression Methodology Are Without Merit

239. Along with in-sample prediction, I also presented two price structure analyses in my Initial Report as evidence of widespread impact to Class Members.³⁹⁹ *First*, I conducted a “common shock” analysis. This involved flagging instances in which a Defendant’s average Effective Institutional Price decreased by at least five percent year-over-year in the actual world, and then analyzing whether each student income decile at said Defendant university also experienced a decrease in average Effective Institutional Prices.⁴⁰⁰ I found that over 85 percent of income decline groups subject to the common shock also experienced a decrease in their average Effective Institutional Price.⁴⁰¹ *Second*, I ran price structure regressions, which involve regressing a Class Member’s Effective Institutional Prices on the average Effective Institutional Prices paid by *other* Class Members.⁴⁰² I found that a Class Member’s Effective Institutional Prices are highly correlated with Effective Institutional Prices paid by other Class Members.⁴⁰³ As explained in my Initial Report, price structure regressions are a common and standard method for proving Common Impact in

399. Singer Report §III.B.2.

400. *Id.* ¶264.

401. *Id.* ¶265. In my reply workpapers, I provide revised results of this common shock analysis using the revised regression data that is adjusted to incorporate all of Dr. Hill’s data-processing adjustments marked as “Accepted” = “Y” outlined in Table 5. I find that 90 percent of income decile groups subject to the common shock also experienced a decrease in their average Effective Institutional Price.

402. *Id.* ¶266.

403. *Id.* ¶267; *id.* Table 14.

antitrust litigation, including in *High-Tech Employee Antitrust* and in *Arizona Travel Nurses*.⁴⁰⁴ I also provided qualitative evidence that Defendants maintain a goal of preserving “horizontal equity” across students—that is, families of similar socioeconomic status and size should receive similar aid.⁴⁰⁵

240. Dr. Stiroh claims that my common-shocks analysis and price structure regressions are both flawed, but she does not acknowledge or dispute the qualitative evidence that I provide regarding Defendants’ horizontal equity considerations.

241. Dr. Stiroh claims that my common-shocks analysis is flawed in part because the use of averages obfuscates individual differences.⁴⁰⁶ Her argument amounts to stating that Class Members within each income decile are subject to different pricing decisions, and so decreases to the average Effective Institutional Price of an income decile do not equate to lower prices for every Class Member subsumed in said decile. My common-shocks analysis is intended to provide corroboratory evidence that a generalized overcharge would impact students across the income distribution. My analysis shows exactly that—both lower, middle-, and higher-income student groups are impacted by a generalized decrease in average Effective Institutional Prices. That there could be a specific instance of a Class Member in a particular decile that does not benefit from a lower Effective

404. See also *In Re Broiler Chicken Grower Class Certification* at 40-41 (“Singer also assesses widespread impact by using an econometric technique known as correlation analysis to determine the existence of a pay structure. This is a standard methodology regularly used in antitrust litigation for demonstrating impact with class-wide evidence.”). I was the plaintiffs’ economic expert in *Arizona Travel Nurses*. The district court accepted my methodology for proving antitrust impact in *Johnson v. Arizona Hospital & Healthcare Ass’n*, Case No. CV 07-1292-PHX-SRB, 2009 WL 5031334 (D. Ariz. Jul. 14, 2009). The same “two-step” methodology utilized in *Johnson* was accepted by the court granting class certification in *In re High-Tech Employees Antitrust Litigation*, 985 F. Supp. 2d 1167, 1206 (N.D. Cal. 2013) (“Plaintiffs noted that Dr. Leamer’s approach followed a roadmap widely accepted in antitrust class actions that uses evidence of general price effects plus evidence of a price structure to conclude that common evidence is capable of showing widespread harm to the class.”). See also, e.g., *Johnson v. Arizona* (finding predominance where conduct was alleged to suppress bill rates for nurses generally and evidence was presented that bill rates were correlated with nurse pay rates).

405. *Id.* ¶269.

406. Stiroh Report ¶203.

Institutional Price is not relevant to this analysis—that is what my in-sample prediction and price structure regressions are for.

242. Dr. Stiroh also argues that what I define as common shocks do not inform whether Class Members would pay lower prices in a but-for world.⁴⁰⁷ I disagree. I define a common shock as being a five percent reduction in a Defendant’s average Effective Institutional Price. Contrary to Dr. Stiroh’s claim, this is one of the most directly informative scenarios for determining whether Class Members would pay lower Effective Institutional Prices in a but-for world. A removal of the Challenged Conduct would be expected to result in a non-trivial reduction (i.e., “shock”) in Effective Institutional Prices as shown by my overcharge regressions. My common-shock analysis analyzes this same scenario using instances where there was a common and non-trivial Effective Institutional Price reduction.

243. Dr. Stiroh asserts that my price structure regressions, which showed that changes in Effective Institutional Prices are highly correlated, provide no insight as to whether all or nearly all Class Members were impacted.⁴⁰⁸ Dr. Stiroh illustrates a hypothetical example where market-wide factors cause prices for all customers to increase by two percent over time, but prices for only a subset of customers increase by ten percent due to an additional overcharge.⁴⁰⁹ She asserts that my methodology is insufficient to establish impact because it would report a price correlation between both groups, even though only one group was subject to an overcharge.⁴¹⁰

244. Dr. Stiroh’s example ignores how my price structure regressions function. I included

407. *Id.* ¶205.

408. *Id.* ¶203. In my reply workpapers, I provide revised results of my price structure regressions using the revised regression data that is adjusted to incorporate all of Dr. Hill’s data-processing adjustments marked as “Accepted” = “Y” outlined in Table 5. I continue to find a high degree of correlation between a Class Member’s Effective Institutional Price and the average Effective Institutional Prices of their peers, consistent with a pricing structure.

409. *Id.* ¶204.

410. *Id.* ¶204.

numerous control variables which account for common, market-wide factors across students, including a time trend, lagged excess endowment investment returns, lagged institutional tuition revenues per full-time equivalent undergraduate, the percent of first-year full-time equivalent undergraduates receiving financial aid, a one year-lag of unemployment, a COVID dummy variable, and real GDP.⁴¹¹ In Dr. Stiroh's hypothetical scenario, these control variables would account for the general two percent increase in Effective Institutional Prices. The regression can be thought of as "subtracting out" this market-wide variation. For instance, applying her example to my regression, the regression would (effectively) subtract out the common two percent increase for all students, and compare only the remaining student-specific price variation. Put differently, the regression would compare the student-specific eight percent overcharge for the student group impacted by the overcharge to a student-specific zero percent overcharge for the other student group. Therefore, contrary to Dr. Stiroh's flawed illustration, my price structure regressions would not find correlation in Effective Institutional Prices between students in each group if there were none.

C. Defendants' Experts Do Not Identify Any Procompetitive Benefits to Justify the Existence of the 568 Group

245. Dr. Hill concludes that "there are no anticompetitive effects that would outweigh any procompetitive benefits of the 568 Group, as identified by Dr. Long in her report."⁴¹² Given that Dr. Hill does not describe any procompetitive benefit of the 568 Group himself, my response to Dr. Long's proposed procompetitive benefits addresses Dr. Hill as well.

246. Dr. Stiroh posits generally that "benchmarking can benefit students."⁴¹³ However, to the extent that schools engaged in benchmarking before, during, and after the conspiracy, benchmarking could not be a procompetitive justification attributable to the conspiracy. Indeed, Dr.

411. Singer Report Table 14.

412. Hill Report ¶14.

413. Stiroh Report ¶102.

Stiroh only offers deposition statements from Defendants themselves as evidence in support of her general proposition. Moreover, Dr. Stiroh's example of Yale only serves to rebut her own theory. As an example of her "benefits" to students, Dr. Stiroh explains:

Likewise, in the 2008–2009 academic year Yale eliminated parent contributions for families making less than \$60,000 and limited all other parent contributions to between one and 15 percent of total income in response to Harvard's "Ten Percent Plan" in which Harvard was advertising a family contribution capped at 10 percent of family income.⁴¹⁴

247. However, the change that Dr. Stiroh describes occurred *after Yale left the conspiracy*. Yale acknowledged that it left the 568 Group precisely because membership in the group restrained its financial aid generosity.⁴¹⁵ In attempting to claim the existence of procompetitive justifications, Dr. Stiroh only manages to underscore the anticompetitive effects of the alleged conspiracy.

248. Dr. Stiroh also argues that "Numerous Defendants specifically implemented no-loan policies throughout the Class Period."⁴¹⁶ I note again that Yale increased its generosity, including adopting its "no-loan" plan in 2008-2009, *after leaving the alleged conspiracy*.⁴¹⁷ Likewise, Rice and Chicago both adopted no-loan policies in the year after exiting the 568 Group.⁴¹⁸ Brown and Emory left the conspiracy in 2013 and adopted no-loan policies in 2017 and 2022, respectively.⁴¹⁹ Thus, to the extent that Dr. Stiroh implies that no-loan policies represent some procompetitive justification, not only does she not offer any attempt to quantify that benefit, but the evidence indicates the

414. Stiroh Report ¶102.

415. Caitlin Roman, *University Leaves Financial Aid Group*, YALE DAILY NEWS (Sept. 26, 2008), <https://yaledailynews.com/blog/2008/09/26/university-leaves-financial-aid-group/>.

416. Stiroh Report ¶104.

417. Jon Victor, *Despite Perkins expiration, little impact on campus*, YALE NEWS, (Oct. 9, 2015), <https://yaledailynews.com/blog/2015/10/09/despite-perkins-expiration-little-impact-on-campus>.

418. Long Report Figure 9 (showing Chicago enacted a no-loan policy in the 2015 academic year. My Class Period definition counts Chicago as leaving the 568 Group starting in the 2015 academic year); Stiroh Report ¶140 (stating that Rice no longer had institutional loans starting in 2022. My Class Period definition counts Rice as leaving the 568 Group starting in 2022.).

419. Natalie Villacres, *Brown makes elimination of undergraduate loans permanent*, BROWN DAILY HERALD, (Mar. 23, 2023), <https://www.browndailyherald.com/article/2023/03/brown-makes-elimination-of-undergraduate-loans-permanent>. Stiroh Report ¶140 (stating that Emory began a no-loan policy in 2022. My Class Period definition counts Emory as having left the 568 Group starting in 2013).

opposite. Defendant schools increased their financial aid generosity after departing the alleged conspiracy.

249. Dr. Long and Dr. Stiroh both claim that some Class Members could benefit from the CM given that particular alterations that the CM provides can be beneficial to some students. I address this argument in section II.B above. My EFC regressions in Table 1 and my in-sample prediction analysis from my Initial Report together indicate that the EFCs resulting from the Challenged Conduct were restricted and that all or nearly all Class Members paid higher Effective Institutional Prices as a result of the Challenged Conduct. These results indicate that specific instances of the CM EFC being lower than the Base IM EFC are not valid procompetitive benefits, given that all or nearly all Class Members still had higher Effective Institutional Prices and likely had higher EFCs than in the but-for world as a result of the alleged conspiracy.

250. Dr. Stiroh similarly argues that procompetitive benefits relate to the use of COFHE colorbooks, which she argues Defendants use for benchmarking in a way that can benefit students.⁴²⁰ The use of COFHE data is only one mechanism of sharing CSI that is established by the 568 Group. As I described in my Initial Report, Defendants also completed surveys and attended meetings as two alternative methods of CSI sharing.⁴²¹ Dr. Stiroh does not explain how these methods could be considered procompetitive, nor does she make any attempt to quantify any such benefits. Furthermore, as Dr. Long argued, COFHE data was used by both Defendants and non-Defendants prior to, during, and after the 568 Group was formed.⁴²² Therefore, the use of COFHE data for

420. Stiroh Report ¶¶101-102.

421. *See, e.g.*, Singer Report §II.A.2 f.

422. Long Report ¶242. Note, however, that while there is not perfect overlap between the COFHE members and the 568 Group, the use of COFHE data did contribute to the 568 Group sharing CSI. That is to say, while it is one method employed by the 568 Group, it is not the only method of sharing CSI implemented by the 568 Group under the alleged conspiracy. *See* Section II.B, *supra*, for additional discussion.

benchmarking could not be considered a procompetitive benefit resulting from the alleged conspiracy.

251. Dr. Long asserts that “Defendants are not revenue-maximizing like for-profit businesses.”⁴²³ For one thing, Defendants favored students from wealthier families over students from non-wealthy families. This is consistent with revenue maximization—admitting students from wealthier families at higher rates than other students is more likely to result in greater donations, which grow Defendants’ endowments. In Appendix 4, I test Dr. Long’s hypothesis that Defendants did not revenue maximize by comparing the admissions rates for students that were “priority-designated” to other, non-priority designated students at multiple Defendant universities that had produced data allowing me to distinguish between these two student types. For Cornell, Georgetown, MIT, Notre Dame, and Penn, I find that priority-designated students were admitted at higher rates than other, non-priority designated students, and that these differences in admissions rates are both economically and statistically significant. Yet, admitted priority-designated students at these Defendant universities generally have ACT and SAT scores equal to or less than the scores for non-priority designated admits. These findings combined indicate that priority-designated students tend to be admitted at higher rates, but with worse standardized test scores than non-priority designated students on average. This finding is inconsistent with Dr. Long’s claim that Defendants do not revenue maximize.

252. Therefore, no Defendant Expert presents a valid argument for why the Challenged Conduct would facilitate procompetitive benefits. Even if such benefits existed, the general anticompetitive effect of the Challenged Conduct outweighed any such benefits, as evidenced by my

423. Long Report ¶22.

regression analyses.⁴²⁴ Moreover, if such procompetitive benefits existed, Yale's explanation for leaving the 568 Group (stating that it was necessary to offer more financial aid) makes no sense.⁴²⁵

IV. AGGREGATE DAMAGES

253. No expert in this case disputes my methodology for calculating aggregate damages to the Class using my overcharge regression model and "but-for world" framework I developed in my Initial Report.⁴²⁶ As explained in Part III.A.2.c, I incorporate a handful of Defendants' critiques pertaining to data processing and control variable construction that have some merit. These adjustments alter my calculation of aggregate damages in two ways.

254. *First*, the conduct coefficient of my primary regression model, which measures the artificial overcharge in Effective Institutional Prices per Class Member and academic year resulting from the Challenged Conduct, decreases from 1,497 in my Initial Report to 1,202 in this rebuttal report, as shown in Table 6 above.

255. *Second*, the data-processing adjustments that I accept decreases the count of Class Member-academic years from approximately 579,000 thousand to 570,000. This decrease in my count of Class Member-academic years is primarily due to my adjustment of the Class Period for Duke and Rice. Under Counsel's instruction, I accept Dr. Hill's updated Class Period definition to

424. Singer Report Table 11. *See also* Table 1, *supra*.

425. Caitlin Roman, *University Leaves Financial Aid Group*, YALE DAILY NEWS (Sept. 26, 2008), <https://yaledailynews.com/blog/2008/09/26/university-leaves-financial-aid-group/>.

426. Singer Report ¶267.

-152-

no longer count Duke or Rice as participating in the Challenged Conduct in 2022, and to no longer count Rice as participating in 2012–2014.⁴²⁷

256. In Table 7, I present updated aggregate damages using the artificial overcharge obtained in my updated Effective Institutional Price regressions. I find that Class Members suffered aggregate damages of \$685 million over the Class Period in the form of artificially inflated Effective Institutional Prices.

TABLE 7: UPDATED AGGREGATE DAMAGES (\$ MILLIONS)

Number of Class Member-academic years subject to the Challenged Conduct	570,176
Effective Institutional Price overcharge per Class Member-academic year resulting from the Challenged Conduct	\$1,202
Aggregate Damages (\$ millions)	\$685.3

Notes: Effective Institutional Price overcharge comes from the conduct coefficient in Table 6, column 6. Damages calculation uses revised data adjusted to incorporate all of Dr. Hill’s data-processing adjustments marked as “Accepted” = “Y” outlined in Table 5.

* * *

Hal J. Singer, Ph.D.:



Executed on October 7, 2024

427. Along with the change in Class Period, Dr. Hill’s data-processing adjustments also result in changes to Class Member counts by academic year for Defendants for which Dr. Hill adjusts the academic year. For instance, Dr. Hill claims that Brown’s aid-year variable refers to the spring semester, rather than the fall semester, and that Brown’s data should therefore be coded based on their aid year minus one. Hill Report ¶193; Point 1; Bullet 2. Dr. Hill incorporates awards classification crosswalks for certain Defendants, which affects my calculation of their institutional grant aid. This alters my count of Class Member-academic years because the adjustment of institutional grant aid can result in these students counting as in-or-out of the Class based on whether they are considered a “full-ride” or zero institutional grant aid recipient.

APPENDIX 1: MATERIALS RELIED UPON

BATES DOCUMENTS AND DEPOSITION EXHIBITS

BROWN_0000002706

BROWN_0000009178

COFHE-02-00006617

COFHE-02-00011341

Columbia_00005800

Columbia_00016697

Columbia_00056363

Columbia_00058983

Columbia_00179751

Columbia_00277826

Columbia_00298393

CORNELL_LIT0000002448

CORNELL_LIT0000004362

CORNELL_LIT0000022406

CORNELL_LIT0000038218

CORNELL_LIT0000100841

CORNELL_LIT0000112401

CORNELL_LIT0000214653

CORNELL_LIT0000252669

CORNELL_LIT0000272526

-154-

CORNELL_LIT0000351496

DARTMOUTH_0000158513

DARTMOUTH_0000359371

DARTMOUTH_0000359527

Deposition of John DeGioia (Feb. 16, 2024) Exhibit 13

Deposition of Patricia McWade (Dec. 8, 2023) Exhibit 3

DUKE568_0081152

DUKE568_0098047

DUKE568_0124311

DUKE568_0155589

Emory_568Lit_0000577

Emory_568Lit_0006213

Emory_568Lit_0006214

Emory_568Lit_0058886

GTWNU_0000009789

GTWNU_0000039045

GTWNU_0000071253

GTWNU_0000079934

GTWNU_0000107168

GTWNU_0000193182

GTWNU_0000317271

GTWNU_0000347907

-155-

GTWNU_19960

Hill Workpapers, Summary of data processing updates.xlsx

MIDDLEBURY001293

MITLIT-000074274

MITLIT-000078586

MITLIT-000169873

ND_0006828

ND_0006830

ND_0019093

ND_0065243

ND_0158341

ND_STRUCTURED_000012

ND_STRUCTURED_000014

PENN568-LIT-00000002

PENN568-LIT-00004598

PENN568-LIT-00018593

PENN568-LIT-00029001

PENN568-LIT-00069194

PENN568-LIT-00089925

PENN568-LIT-00133526

PENN568-LIT-00137113

PENN568-LIT-00167414

-156-

RICE_LIT00000007155

RICE_LIT00000007623

RICE_LIT00000033957

RICE_LIT00000067255

UCHICAGO_STRUCTURED_1000000025

UCHICAGO_STRUCTURED_1000000026

UCHICAGO-0000052966

UID_0000404135288

VANDERBILT-00000528

Vanderbilt-00041936

VANDERBILT-00047310

Vanderbilt-00204227

VANDERBILT-00285876

Vanderbilt-00389088

YALE_LIT_0000013965

DEPOSITION TRANSCRIPTS

Deposition of Anne Walker (Sept. 8, 2023)

Deposition of Brent Tener (Jul. 12, 2023)

Deposition of Chris Watson (Oct. 10, 2023)

Deposition of Diane Corbett (Jan. 8, 2024)

Deposition of Edward Malloy (Sept. 7, 2023)

Deposition of Elaine Varas (Aug. 2, 2023)

Deposition of James Tilton (May 15, 2023)

Deposition of Jason Locke (Dec. 11, 2023)

Deposition of Jessica Marinaccio (Sept. 18, 2023)

Deposition of John DeGioia (Feb. 16, 2024)

Deposition of John Gaines (Aug. 3, 2023)

Deposition of Karen Cooper (Mar. 27, 2024)

Deposition of Karl Furstenberg (Jul. 14, 2023)

Deposition of Malina Chang (Sept. 13, 2023)

Deposition of Mary Nucciarone (Nov. 1, 2023)

Deposition of Michael Hall (Sept. 13, 2023)

Deposition of Miranda McCall (Aug. 29, 2023)

Deposition of Patricia McWade (Dec. 8, 2023)

Deposition of Peter Wyatt (Feb. 15, 2024)

Deposition of Scott Wallace-Juedes (July 7, 2023)

Deposition of Thomas McDermott (Aug. 18, 2023)

LITERATURE

Adam Smith, *An Inquiry into the Wealth of Nations* (MetaLibri 2007)

Alberto Abadie, Guido Imbens, Susan Athey, and Jeffery Wooldridge, *When Should You Adjust Standard Errors for Clustering?*, 138(1) *Quarterly Journal of Economics* 1-35 (2023)

American Bar Association, *Econometrics: Legal, Practical, and Technical Issues* (ABA Book Publishing 2nd ed. 2014)

American Bar Association, *Proving Antitrust Damages: Legal and Economic Issues* (ABA Book Publishing 3rd ed. 2017)

Andrew C. Baker, David F. Larcker, and Charles C.Y. Wang, *How Much Should We Trust Staggered Difference-In-Differences Estimates*, 144(2) *Journal of Financial Economics* 370-395 (2022)

Andrew Gillen and Jonathan Robe, *Stop Misusing Higher Education Specific Price Indices*, Center for College Affordability and Productivity (2011)

Andrew Goodman-Bacon, Difference-In-Differences with Variation in Treatment Timing, 225(2) *Journal of Econometrics* 254-277 (2021)

Andrew J. Leone, Miguel Minutti-Meza, and Charles E. Wasley, Influential Observations and Inference in Accounting Research, 94(6) *The Accounting Review* 337-364 (2019)

Bart Meuleman, Geert Loosveldt, and Viktor Emonds, *Regression Analysis: Assumptions and Diagnostics*, in *Sage Handbook of Regression and Causal Inference* (Henning Best and Christof Wolf eds. 2015)

Bendix Carstensen, *Do not group quantitative variables*, in *Epidemiology with R* (Oxford University Press 2020)

Brantly Callaway and Pedro H.C. Sant'Anna, *Difference-in-Differences with Multiple Time Periods*, 225(2) *Journal of Econometrics* 200-230 (2021)

Bruce E. Hansen, *The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity*, 15(4) *Journal of Economic Perspectives* 117-128 (2001)

Carlos Cinelli, Andrew Forney, and Judea Pearl, A Crash Course in Good and Bad Controls, 53(3) *Journal Sociological Methods and Research* 1071-1104 (2024)

Clement de Chaisemartin and Xavier D'Haultfoeuille, *Two-way fixed effects estimators with heterogeneous treatment effects*, 110(9) *American Economic Review* 2964-2996 (2020)

Daniel L. Millimet and Christopher F. Parmeter, *Accounting for Skewed or One-Sided Measurement Error in the Dependent Variable*, IZA Institute of Labor Economics, Discussion Paper No. 12576 (Aug. 2019)

Daniel L. Rubinfeld, *Reference Guide on Multiple Regression*, 3 *Reference Manual on Scientific Evidence* 303-357 (2011)

Daniel Westreich and Sander Greenland, *The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients*, 177(4) *American Journal Epidemiology* 292-298 (2013)

David Belsley, Edwin Kuh, and Roy Welch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* (Wiley Interscience 2004)

David Genesove & Wallace P. Mullin, *Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890-1914*, 29(2) *RAND Journal of Economics* 355-377 (1998)

David McKenzie, *When should you cluster standard errors? New wisdom from the econometrics oracle*, World Bank Blog (Oct. 16, 2017), <https://blogs.worldbank.org/en/impactevaluations/when-should-you-cluster-standard-errors-new-wisdom-econometrics-oracle>

Douglas Altman and Patrick Royston, *The cost of dichotomising continuous variables*, 332 The BMJ 1080(2006)

Douglas Altman, *Categorizing Continuous Variables* (John Wiley & Sons 2005)

H. Peter Boswijk, Maurice J. G. Bun, and Maarten Pieter Schinkel, *Cartel Dating*, 34 Journal of Applied Econometrics 26-42 (2017)

Hal Singer and Kevin Caves, *Applied Econometrics: When Can an Omitted Variable Invalidate a Regression*, 17(3) *The Antitrust Source* (2017)

Information Exchanges between Competitors under Competition Law, OECD Directorate for Financial and Enterprise Affairs Competition Committee: Germany (Jul. 11, 2011), https://www.oecd.org/content/dam/oecd/en/publications/reports/2011/07/information-exchanges-between-competitors-under-competition-law_bd644d8b/327f7dd3-en.pdf

Jacob Cohen, *The Cost of Dichotomization*, 7(3) *Applied Psychological Measurement* 249-253 (1983)

James G. MacKinnon, Morten Ørregaard Nielsen, and Matthew D. Webb, *Cluster-robust inference: A guide to empirical practice*, 232(2) *Journal of Econometrics* 272-299 (2023)

James H. Stock and Mark W. Watson, *Introduction to Econometrics* (Addison Wesley 2nd ed. 2006)

Jeffrey M. Perloff, *Microeconomics* (Pearson 7th ed. 2015)

Jeffrey Wooldridge, *Applied Econometrics, A Modern Approach* (5th ed. 2012)

Joe H. Sullivan, Merrill Warkentin, and Linda Wallace, *So many ways for assessing outliers: What really works and does it matter?*, 132 *Journal of Business Research* 530-543 (2021)

Johannes Odenkirchen, *Pricing behavior in partial cartels*, Düsseldorf Institute for Competition Economics, No. 299 Discussion Paper (Sept. 2018)

John W. Tukey, *A Survey of Sampling from Contaminated Distributions, in Contributions to probability and statistics* in Essays in honor of Harold Hotelling 448-485 (Ingram Olkin et al. eds. Stanford Univ. Press 1960)

Joseph E. Harrington, *Detecting Cartels*, John Hopkins University, Department of Economics, Working Paper No. 526 (2005)

Joseph E. Harrington, Jr., *Post-Cartel Pricing During Litigation*, 52(4) The Journal of Industrial Economics 517-533 (2004)

Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly harmless econometrics: An empiricist's companion* (Princeton University Press 2009)

Judea Pearl, *On a Class of Bias-Amplifying Variables that Endanger Effect Estimates*, Proceedings on Uncertainty in Artificial Intelligence (UAI2010), Technical Report R-356 (Jul. 2010), https://ftp.cs.ucla.edu/pub/stat_ser/r356.pdf

Liyang Sun and Sarah Abraham, *Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects*, 225(2) Journal of Econometrics 175-199 (2021)

Margaret C. Levenstein and Valerie Y. Suslow, *What Determines Cartel Success?*, 44(1) Journal of Economic Literature 43-95 (2006)

N. Gregory Mankiw, *Principles of Microeconomics* (Cengage 8th ed. 2016)

P.E. Pontius, *Measurement Philosophy of the Pilot Program for Mass Calibration*, U.S. Department of Commerce, National Bureau of Standards, Technical Note 288 (May 6, 1966)

Paul Hünermund and Beyers Louw, *On the Nuisance of Control Variables in Causal Regression Analysis*, 0 Organizational Research Methods 1-14 (2023)

Peter Davis and Eliana Garcés, *Quantitative Techniques For Competition and Antitrust Analysis* (Princeton University Press 2010)

R. Wilms, E. Mäthner, L. Winnen, and R. Lanwehr, *Omitted variable bias: A threat to estimating causal relationships*, 5 Methods in Psychology 1-10 (2021)

Richard K. Vedder, *Going Broke by Degree: Why College Costs Too Much*, Washington DC: American Enterprise Institute (2004)

Richard K. Vedder, *Federal Tax Policy Regarding Universities: Endowment and Beyond*, Washington DC: Center for College Affordability and Productivity (2008)

Robert B. Archibald & David H. Feldman, *Explaining Increases in Higher Education Costs*, 79(3) The Journal of Higher Education 268-295 (2008)

Roger D. Blair and Virginia G. Maurer, *Umbrella Pricing and Antitrust Standing: An Economic Analysis*, 4 Utah Law Review 763-796 (1982)

Ronald L. Wasserstein and Nicole A. Lazar, *The ASA's Statement on p-Values: Context, Process, and Purpose*, 70(2) *The American Statistician* 129-133 (2016)

Sander Greenland, Stephen J. Senn, et al., *Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations*, 31 *European Journal of Epidemiology* 337-350 (2016)

Sandy Baum, *A Primer on Economics for Financial Aid Professionals*, College Board and the National Association of Student Financial Aid Administrators 4 (1996)

Seung-Whan Choi, *The effect of outliers on regression analysis: regime type and foreign direct investment*, 4(2) *Quarterly Journal of Political Science* 153-165 (2009)

Ted Tatos, *Relevant Market Definition and Multi-Sided Platforms Post Ohio v. American Express: Evidence from Recent NCAA Antitrust Litigation*, 10(2) *Harvard Journal of Sports and Entertainment Law* 147-172 (2019)

Theon van Dijk & Frank Verboven, *Quantification of Damages*, in 3 *Issues in Competition Law and Policy* 2331-2348 (ABA Section of Antitrust Law 2008)

Thomas E. MaCurdy & John H. Pencavel, *Testing between Competing Models of Wage and Employment Determination in Unionized Markets*, 94(3) *Journal of Political Economy* S3-S39 (Jun. 1986)

Yu Awaya & Vijay Krishna, *Information Exchange in Cartels*, 51(2) *RAND Journal of Economics* 421-446 (2020)

LEGAL DOCUMENTS

In re Broiler Chicken, Case No. 16-cv-8637, 2022 WL 1720468 (N.D. Ill. May 27, 2022)

In re Capacitors Antitrust Litig. (No. III), Case No. No. 17-md-02801, 2018 WL 5980139 (N.D. Cal. Nov. 14, 2018)

In re Domestic Drywall Antitrust Litig., 322 F.R.D. 188 (E.D. Pa. 2017)

In re High-Tech Employees Antitrust Litigation, 985 F. Supp. 2d 1167 (N.D. Cal. 2013)

In re Korean Ramen Antitrust Litig., Case No. 13-cv-04115, 2017 WL 235052 (N.D. Cal. Jan. 19, 2017)

In Re: Broiler Chicken Growing Antitrust Litigation (No. II), Case No. 6:20-MD-02977-RJS-CMR (E.D. Ok May 8, 2024)

Johnson v. Arizona Hospital & Healthcare Ass'n, No. CV 07-1292-PHX-SRB, 2009 WL 5031334 (D. Ariz. Jul. 14, 2009)

Olean Wholesale Grocery Coop., Inc. v. Bumble Bee Foods, LLC, 31 F.4th 651 (9th Cir. 2022)

State of NY v. Facebook, Case No. 21-7078, Brief of Economists as Amici Curiae in Support of Plaintiff-Appellants and Reversal (D.C. Cir. Jan. 28, 2022), <https://www.cohenmilstein.com/wp-content/uploads/2023/07/New-York-v-Facebook-Economists-Amicus-Filed-01282022.pdf>

OTHER PUBLICLY AVAILABLE MATERIALS

Abby Spiller, *Duke rises to No. 6 in U.S. News and World Report national ranking, highest in 19 years*, The Chronicle (Sept. 23, 2024), <https://www.dukechronicle.com/article/2024/09/duke-university-sixth-us-news-and-world-report-national-university-ranking-america-highest-since-2006-top-10-caltech-johns-hopkins-northwestern>

About The Daily Pennsylvanian, Inc., Daily Pennsylvanian, <https://www.thedp.com/page/about> (last visited Oct. 2024)

Alan Blinder, *The U.S. News College Rankings Are Out. Cue the Rage and Obsession*, New York Times (Sept. 24, 2024), <https://www.nytimes.com/2024/09/24/us/us-news-rankings-colleges.html>

Anushka Shorewala, *U.S. News Ranks Cornell No. 11 University in Country, Best in New York*, The Cornell Daily Sun (Sept. 23, 2024), <https://cornellsun.com/2024/09/23/cornell-ranked-no-11-university-in-country-best-in-new-york-in-u-s-news-and-world-report/>

Caitlin Roman, *University Leaves Financial Aid Group*, Yale Daily News (Sept. 26, 2008), <https://yaledailynews.com/blog/2008/09/26/university-leaves-financial-aid-group/>

Campus Solutions 9.2: Financial Aid, Oracle, https://docs.oracle.com/cd/F33383_01/psft/pdf/cs92lsfa-b072020.pdf (last visited Sept. 2024)

Can you perform a log transformation in SPSS?, IBM Support (Apr. 16, 2020), <https://www.ibm.com/support/pages/can-you-perform-log-transformation-spss>

Cate Latimer, *Brown University drops from Top 10 in 2025 U.S. News Ranking*, The Brown Daily Herald (Sept. 24, 2024), <https://www.browndailyherald.com/article/2024/09/brown-university-drops-from-top-10-in-2025-u-s-news-ranking>

Chris Stipes, *Rice in top 20 of US News 'Best Colleges' rankings*, Rice University News and Media Relations (Sept. 24, 2024), <https://news.rice.edu/news/2024/rice-top-20-us-news-best-colleges-rankings-0>

Consortium on Financing Higher Education, MIT, <https://web.mit.edu/cofhe/> (last visited Oct. 6, 2024)

Costs, The University of Chicago

<https://web.archive.org/web/20210301122831/https://financialaid.uchicago.edu/undergraduate/costs> (archived Mar. 1, 2021)

Costs, The University of Chicago,

<https://web.archive.org/web/20220120155943/https://financialaid.uchicago.edu/undergraduate/costs> (archived Jan. 20, 2022)

Dave Bergman, *How to Get into Rutgers University: Admissions Data & Strategies*, College Transitions (Jul. 30, 2024), <https://www.collegetransitions.com/blog/how-to-get-into-rutgers-university/>

Duke to Rescind Planned Undergraduate Tuition Increase, Reduce Fees for 2020-21 Academic Year, Duke Today (Aug. 1, 2020), <https://today.duke.edu/2020/08/duke-rescind-planned-undergraduate-tuition-increase-reduce-fees-2020-21-academic-year>

Duke Trustees Set Tuition, Reappoint President at Quarterly Meeting, Duke Today (Feb. 27, 2021), <https://today.duke.edu/2021/02/duke-trustees-set-tuition-reappoint-president-quarterly-meeting>

Emma Edmund, *Northwestern goes remote for first- and second-year students, reduces Fall Quarter undergraduate tuition by 10 percent*, The Daily Northwestern (Aug. 28, 2020), <https://dailynorthwestern.com/2020/08/28/campus/northwestern-goes-remote-for-first-and-second-year-students-reduces-fall-quarter-undergraduate-tuition-by-10-percent/>

Financial Aid: Merit-Based Scholarships, Yale, <https://finaid.yale.edu/award-letter/financial-aid-terminology/merit-based-scholarships> (last visited Oct. 2024)

Gillian Diebold, *Penn admits a record-low 7.44 percent of applicants to the Class of 2023*, Daily Pennsylvanian (Mar. 28, 2019), <https://www.thedp.com/article/2019/03/penn-acceptance-ivy-league-regular-decision-admissions-class-2023>

How Recessions Impact Household Net Worth, Federal Reserve Bank of St. Louis (Nov. 23, 2020), <https://www.stlouisfed.org/on-the-economy/2020/november/recessions-impact-household-net-worth/>

Jennifer Ma and Matea Pender, *Trends in College Pricing and Student Aid 2023*, College Board (2023), <https://research.collegeboard.org/media/pdf/Trends%20Report%202023%20Updated.pdf>

Jo-Anne Williams Barnes, *Do Nonprofit Organizations Have Profit and Loss Statements?*, JFW ACCOUNTING SERVICES (Oct. 10, 2022)

Johns Hopkins Rises to No. 6 In 'U.S. News' Best Colleges Rankings, Johns Hopkins University Hub (Sept. 24, 2024), <https://hub.jhu.edu/2024/09/24/us-news-best-colleges-rankings-2024/>

Jon Victor, *Despite Perkins expiration, little impact on campus*, Yale News, (Oct. 9, 2015), <https://yaledailynews.com/blog/2015/10/09/despite-perkins-expiration-little-impact-on-campus>

Ka Yee C. Lee & John W. Boyer, *2020-2021 College Tuition, Housing and Fees*, The University of Chicago (Apr. 13, 2020), <https://college.uchicago.edu/2020-2021-college-tuition-housing-and-fees>

Katie Bartlett, *Penn falls to lowest U.S. News ranking since 1997 while Princeton, MIT notch top spots*, The Daily Pennsylvanian (Sept. 10, 2024), <https://www.thedp.com/article/2024/09/penn-princeton-mit-us-news-rankings-drop>

Kevin Kiley, *Sorry, Wrong Numbers*, Inside Higher Ed (Aug. 12, 2012), <https://www.insidehighered.com/news/2012/08/20/emory-misreported-admissions-data-more-decade>

Laura Diamond, *U.S. News names Emory among top national universities*, Emory News Center (Sept. 24, 2024), https://news.emory.edu/stories/2024/09/er_us_news_undergraduate_rankings_24-09-2024/story.html

Marc Bellemare, *Metrics Monday: What to Do Instead of $\log(x + 1)$* , MarcFBellemare.com (Feb. 26, 2018), <https://marcfbellemare.com/wordpress/12856>

Margaret C. Levenstein and Valerie Y. Suslow, *Cartel bargaining and monitoring: The role of information sharing*, in *The Pros and Cons of Information Sharing*, Swedish Competition Authority (2006)

Memorandum of Understanding, The 568 Presidents' Group, <https://web.archive.org/web/20050311231218/http://568group.org/docs/memo.pdf> (archived 11 Mar. 2005)

Mia Streitberger, *Georgetown Remains at No. 22 in U.S. News 2024 Top Colleges List*, The Hoya (Sept. 23, 2024), <https://thehoya.com/news/georgetown-remains-at-no-22-in-u-s-news-2024-top-colleges-list/>

Michael Thaddeus, *College rankings whistleblower: Exposing inaccurate data was unpleasant but necessary*, CNN (Sept. 22, 2022), <https://www.cnn.com/2022/09/22/opinions/columbia-ranking-inaccurate-data-thaddeus/index.html>

Michelle N. Amponsah and Emma H. Haidar, *84% of Admits Accept Spots in Harvard College Class of 2027*, The Harvard Crimson (May 20, 2023), <https://www.thecrimson.com/article/2023/5/20/class-of-2027-yield-data/>

MIT named No. 2 university by U.S. News for 2024-25, MIT News (Sept. 24, 2024), <https://news.mit.edu/2024/mit-named-no-2-university-us-news-0924>

Muhammad Al Amin, *Panel Data Using Stata: Fixed Effects and Random Effects*, Princeton University Library Research Guides, <https://libguides.princeton.edu/stata-panel-fe-re>

Natalie Villacres, *Brown makes elimination of undergraduate loans permanent*, Brown Daily Herald, (Mar. 23, 2023), <https://www.browndailyherald.com/article/2023/03/brown-makes-elimination-of-undergraduate-loans-permanent>

Nineth Kanieski Koso, *Northwestern reaches historic high at 6th place in U.S. News college rankings*, The Daily Northwestern (Sept. 24, 2024), <https://dailynorthwestern.com/2024/09/24/campus/northwestern-reaches-historic-high-at-6th-place-in-u-s-news-college-rankings/>

PowerFAIDS, College Board, <https://powerfaids.collegeboard.org/media/pdf/powerfaids-financial-aid-management.pdf> (last visited Sept. 2024)

Price Fixing, Bid Rigging, and Market Allocation Schemes: What They are and What to Look For, U.S. Department of Justice (revised Feb. 2021), <https://www.justice.gov/d9/pages/attachments/2016/01/05/211578.pdf>.

Reply by Carlo Lazzaro, *Fixed Effects Regression and the role of Singleton Observations in a Balanced Panel*, STATALIST (Jan. 31, 2023 at 10:08), <https://www.statalist.org/forums/forum/general-stata-discussion/general/1656468-fixed-effects-regression-and-the-role-of-singleton-observations-in-a-balanced-panel>

Sara Levine, *Admission Rates Drop at Penn, Ivies*, Daily Pennsylvanian (Apr. 11, 2003), https://www.thedp.com/article/2003/04/admission_rates_drop_at_penn_ivies

Seth M. Freedman, Alex Hollingsworth, Kosali I. Simon, Coady Wing, and Madeline Yozwiak, *Designing Difference in Difference Studies with Staggered Treatment Adoption: Key Concepts and Practical Guidelines*, National Bureau of Economic Research, NBER Working Paper No. 31842 (November 2023)

Spencer Davis, *Columbia ranks No. 13 in U.S. News rankings, falling one spot*, Columbia Spectator (Sept. 24, 2024), <https://www.columbiaspectator.com/news/2024/09/25/columbia-ranks-no-13-in-us-news-rankings-falling-one-spot/>

The EFC Formula, 2014-2015, U.S. Department of Education, Federal Student Aid, <https://fsapartners.ed.gov/sites/default/files/attachments/efcformulaguide/091913EFCFormulaGuide1415.pdf> (last visited Oct. 2024)

The Table 2 Fallacy, Daggity.net, <https://dagitty.net/learn/graphs/table2-fallacy.html> (last visited Oct. 2024)

What PowerFAIDS Helps You Do, College Board, <https://powerfaids.collegeboard.org/about-powerfaids/what-powerfaids-helps-you-do> (last visited Sept. 2024)

Yale Cuts Costs for Families and Students, YaleNews (Jan. 14, 2008), <https://news.yale.edu/2008/01/14/yale-cuts-costs-families-and-students>

Yunkyo Kim, *Northwestern to increase total cost by 3.6 percent, financial aid by more than 8 percent*, The Daily Northwestern (Jun. 14, 2021), <https://dailynorthwestern.com/2021/06/14/campus/northwestern-to-increase-total-cost-by-3-6-percent-financial-aid-by-more-than-8-percent/>

CASE MATERIALS

Brown's Responses to Questions re: Structured Data (Jan. 26, 2024)

Expert Report of Bridget Terry Long, Ph.D. (Aug. 7, 2024)

Expert Report of George Bulman, Ph.D. (May 14, 2024)

Expert Report of Lauren J. Stiroh, Ph.D. (Aug. 7, 2024)

Expert Report of Nicholas Hill, Ph.D (Aug. 7, 2024)

Georgetown's August 8, 2023, Structured Data Responses

Georgetown's Response Nos. 2-3 to Plaintiffs' First Set of Requests for Admission (Nov. 13, 2023)

Henry et al. v. Brown University et al., Errata I Expert Report of Hal J. Singer, Ph.D., May 28, 2024

Henry et al. v. Brown University et al., Errata II Expert Report of Hal J. Singer, Ph.D., June 10, 2024

Henry et al. v. Brown University et al., Expert Report of Hal J. Singer, Ph.D., May 14, 2024.

Plaintiff Sia Henry's Supplemental Responses and Objections to Defendants' First Set of Interrogatories, *Henry v. Brown University*, Case No. 1:22-cv-00125, (N.D. Ill. Jun. 27, 2023)

Report of David L. Yermack (Aug. 7, 2024)

Request No. 7, Plaintiffs' First Set of Requests for Production of Documents, *Henry v. Brown University*, Case No. 1:22-cv-00125 (N.D. Ill. Sept. 10, 2022)

-167-

PRIVILEGED AND CONFIDENTIAL
PREPARED FOR COUNSEL

APPENDIX 2: DR. HILL'S QUALITATIVE EVIDENCE REGARDING DEFENDANTS' DIFFERING CALCULATIONS APPEARS TAKEN OUT OF CONTEXT AND IS INCONSISTENT WITH HIS CONCLUSIONS

257. Dr. Hill argues that Defendants calculate EFCs and net prices in different manners from each other, indicating a lack of support for the alleged conspiracy.⁴²⁸ To support his findings, Dr. Hill identified limited qualitative evidence, mostly in the form of backward-looking testimony, that he contends is consistent with Defendants having different calculations for EFCs and net prices.⁴²⁹ As explained earlier, Dr. Hill's consideration of the qualitative evidence does not even attempt to address the substantial qualitative evidence I previously cited that is inconsistent with his conclusions and supports the claim that the Challenged Conduct would have resulted in artificially inflated Effective Institutional Prices.⁴³⁰ Continuing to review such evidence through an economic lens, having reviewed Dr. Hill's qualitative evidence, it appears to lack context, and, for the reasons below, the qualitative evidence remains consistent with Defendants engaging in the Challenged Conduct that resulted in artificially inflated Effective Institutional Prices.

A. Qualitative Evidence Does Not Support That Defendants Calculated Family Contributions Differently

258. Dr. Hill cites the 568 Group's CM guidelines and Dr. Long's report to conclude that the guidelines did not "require[] schools to use these elements in a particular way" and were "free to adopt any suggested improvements" or "ignore any that they did not."⁴³¹ Yet contemporaneous qualitative evidence is inconsistent with this understanding. A stated "purpose" of the 568 Group to which members formally ascribed was to "agree upon common principles of analysis for determining the financial need of undergraduate financial aid applicants."⁴³² Similarly, a stated "Value of 568

428. Hill Report §5.

429. *Id.* ¶¶112-14, ¶¶131-32.

430. Singer Report Appendix 7.

431. Hill Report ¶113.

432. CORNELL_LIT0000002448.

Membership” document echoed the common refrain found in the consensus methodology guidelines and early 568 Group discussions that the group was focused on the “consistent treatment of families in similar circumstances across member institutions” and “diminishing or eliminating divergent results” in need analysis.⁴³³ To the extent there was divergence in needs analysis, members typically understood that they could not calculate a family contribution to be *lower* than what the consensus methodology provided.⁴³⁴

259. Dr. Hill cites a 2015 survey of 568 Group members to argue that “at least some of the responding schools reported deviating from the consensus methodology guidelines.”⁴³⁵ But Dr. Hill omits that the survey found substantial compliance with the methodology. The same survey showed that ninety percent of the responding schools agreed with all of the six Core Principles related to financial aid awards.⁴³⁶ In addition, for the majority of CM elements, the survey responses indicated that only ten percent or fewer of the schools deviated from the CM guidelines. For two of the elements, the responses indicated that twenty percent or fewer of the schools deviated from the CM guidelines, and for two other elements, the responses indicated that the percentage of schools that deviated from the guidelines ranged from slightly above twenty-five percent to slightly above thirty percent.⁴³⁷

260. Dr. Hill cites testimony from Brown’s Former Dean of Financial Aid stating that “[w]e did not use the consensus methodology when calculating [Brown’s] institutional methodology.”⁴³⁸ But he also testified that “[w]hen we talk about the principles of how to determine a family’s ability to pay, general consensus about what components should be used . . . in general, people agreed. I

433. YALE_LIT_0000013965.

434. *See, e.g.*, DARTMOUTH_0000359371 at -379; GTWNU_0000193182.

435. Hill Report ¶114.

436. CORNELL_LIT0000252669 at -672-77.

437. *See generally id.*

438. Hill Report ¶114 (citing Deposition of James Tilton (May 15, 2023) [hereafter Tilton Dep.] 59:13–19).

don't know anyone who outwardly didn't agree."⁴³⁹ In addition, other qualitative evidence reveal that contemporaneously, Brown viewed itself as following the CM, even if, as Dr. Hill notes, it signed a memorandum of understanding that it was not "committing or binding itself" to any particular calculation.⁴⁴⁰

261. Dr. Hill cites the testimony of Caltech's Director of Student Financial Aid suggesting that Caltech did not change any of its need analysis components after joining the 568 Group.⁴⁴¹ But Caltech did not change its need analysis components because it was already following many CM components. Caltech's Director of Student Financial Aid testified that many policies were in accord with the CM, such as (1) requiring information from non-custodial parents; (2) adding back depreciation; (3) using the IM's cost-of-living adjustment; (4) using the standard IM allowance for medical expenses; (5) excluding one-time income on a case-by-case basis; among others.⁴⁴² And she also testified that Caltech adopted most of the 568 core principles.⁴⁴³ In addition, contemporaneous qualitative evidence is consistent with Caltech looking to the CM when fashioning its financial aid.⁴⁴⁴

262. Dr. Hill cites the testimony of Columbia's Financial Aid Director that Columbia was not "always following what the consensus methodology said to the letter, anyway."⁴⁴⁵ But Columbia's Dean of Admissions and Financial Aid testified that "You have a choice as to whether or not . . . to adopt the consensus approach, and we have adopted it."⁴⁴⁶ In addition, Columbia stated in its March 2022 response to the DOJ's Request for Information that "Columbia College and SEAS

439. Tilton Dep. 79:23-80:5.

440. *Compare* Hill Report ¶114, n. 153 with BROWN_0000002706 (showing Brown decided in March 2007 to "[u]se IM/568 methodology" for financial aid awards) and BROWN_0000009178 (concluding in 2008 that "Columbia also uses the 568 Principles for Needs Analysis as do we, so our calculated parent contributions should be similar.").

441. Hill Report ¶114.

442. Deposition of Malina Chang (Sept. 13, 2023) 289:24-300:16.

443. *Id.* 227:9-234:11.

444. COFHE-02-00006617-6618 (showing that Caltech queried the 568 Group members about their approach to home equity).

445. Hill Report ¶114 (citing Deposition of Michael Hall (Sept. 13, 2023) 54:12-24).

446. Deposition of Jessica Marinaccio (Sept. 18, 2023) 375:2-5.

have adopted the Consensus Approach as part of their methodology for calculating a family's parental contribution to the cost of attendance."⁴⁴⁷

263. Dr. Hill cites the testimony of Cornell's Executive Director of Financial Aid that she understood the CM guidelines were "best practices that we at Cornell could choose to adopt or not as we developed our own methodology."⁴⁴⁸ But only a few moments later, she also had no basis to disagree with the statements in the value of 568 Membership, including that "the 568 Group is focused on consistent treatment of families in similar circumstances across member institutions, diminishing or eliminating divergent results."⁴⁴⁹ Furthermore, contemporaneous qualitative evidence reflects that Cornell consistently applied the CM. In 2008, Cornell reported to Congress that a "family's contribution toward educational costs is determined by using the 568 Presidents' Group Consensus Approach methodology."⁴⁵⁰ In 2012, Cornell stated that it "uses the College Board's Institutional Methodology, modified by adjustments agreed on by the 568 Presidents' Group of schools."⁴⁵¹ And in 2019, Cornell procedures stated that it "use[s] both the Consensus Approach (CA) and Federal Methodology (FM) formulas to calculate the Parent and Student Contributions."⁴⁵²

264. Dr. Hill cites the testimony of Dartmouth's Dean of Admissions and Financial Aid, who testified that he was not "aware" whether Dartmouth adopted *any* part of the CM.⁴⁵³ But he also testified that he did not handle financial aid and did not even know how the CM compared to Dartmouth's approach, or if Dartmouth "phase[d] in the consensus methodology."⁴⁵⁴ In contrast, contemporaneous qualitative evidence indicates that Dartmouth understood itself to be a

447. Columbia_00298393 at -398.

448. Hill Report ¶114 (citing Deposition of Diane Corbett (Jan. 8, 2024) [hereafter Corbett Dep.] 110:16-111:3).

449. Corbett Dep. 111:12-15; CORNELL_LIT0000022406.

450. CORNELL_LIT0000100841 at -847.

451. CORNELL_LIT0000351496 at -506.

452. CORNELL_LIT0000272526.

453. Hill Report ¶114 (citing Deposition of Karl Furstenberg (Jul. 14, 2023) [hereafter Furstenberg Dep.] 97:15-16).

454. Furstenberg Dep. 141:14-20, 214:2-4, 221:15-18, 219:18-24.

“participating school” in the 568 Group and that it had to “commit to using the results of the [consensus approach] as the *lowest contribution* that would be expected of a family,” among other requirements.⁴⁵⁵

265. Dr. Hill cites the testimony of Duke's Director of Financial Aid stating that she had "no idea how the consensus methodology compared to" Duke's institutional methodology.⁴⁵⁶ But in that same deposition, she testified that Duke followed the CM on many components, such as collecting noncustodial parent information in cases of divorce; making cost of living adjustments with the CSS tables; capping home equity at 1.2 times total income; applying the IM formula for multiple siblings in college; making allowances for private school tuition; among others.⁴⁵⁷ ■

⁴⁵⁸ In 2008,

459

266. Dr. Hill cites the testimony of Johns Hopkins’s Associate Vice President of Financial Aid that he recalled at one meeting that someone asked “for a show of hands of how many institutions were, in fact, using CM as it was defined. And my recollection was that no one raised their hand.”⁴⁶⁰ But Johns Hopkins itself followed most of the CM and adopted more of it over time as it transitioned

455. DARTMOUTH_0000359371 at -379 (emphasis added).

456. Hill Report ¶114 (citing Deposition of Miranda McCall (Aug. 29, 2023) [hereafter McCall Dep.] 76:13-16).

457. McCall Dep. 217:18-245:13.

458. DUKE568 0081152 at -395.

459. ND_0006828 (emphasis and ellipsis in original).

460. Hill Report ¶114 (citing Deposition of Thomas McDermott (Aug. 18, 2023) [hereafter McDermott Dep.] 90:23-91:14).

from a guest to a formal member,⁴⁶¹ and the anecdote is inconsistent with the extensive qualitative evidence I cited in my Initial Report and herein.

267. Dr. Hill cites the testimony of MIT's Director of Student Financial Services that "[T]here was no agreement to use anything. We could use what we wanted and not use what we did not want to use."⁴⁶² But contemporaneous qualitative evidence shows that MIT understood that the 568 Group "provides MIT financial aid staff with professional development opportunities for jointly discussing and agreeing on principles of need analysis."⁴⁶³ In particular, to the extent MIT did not need to agree, it understood that it could be "more stringent in how it defines financial need, but not more generous."⁴⁶⁴

268. Dr. Hill cites the testimony of Notre Dame's Dean of Financial Aid that she "did not interpret anything about those 568 guidelines as mandatory," and each school "decided which of the tools that were in that manual were appropriate to our institution."⁴⁶⁵ But she agreed that Notre Dame's institutional methodology was based on the Base IM, and it aligned with most components of the CM.⁴⁶⁶ In addition, contemporaneous qualitative evidence shows that Notre Dame recognized that the Consensus Agreement established "certain agreed upon parameters" for awarding financial aid.⁴⁶⁷

269. Dr. Hill cites the testimony of Penn's Director of Financial Aid that "we were not following all of the consensus methodology" during her participation in 568 Group activities.⁴⁶⁸ But during her tenure and before Penn exited the 568 Group on her recommendation, Penn made policy

461. McDermott Dep. 178:3-195:1; Deposition of Peter Wyatt (Feb. 15, 2024) 51:18-52:2; 192:2-14.

462. Hill Report ¶114 (citing Deposition of Leslie Bridson (July 13, 2023) 49:9-15).

463. MITLIT-000074274 at -274 (emphasis added).

464. *Id.*

465. Hill Report ¶114 (citing Deposition of Mary Nucciarone (Nov. 1, 2023) 325:17-326:14).

466. Rule 30(b)(6) Deposition of Mary Nucciarone (Notre Dame) (Apr. 3, 2024) 304:19-22, 230:15-20, 316:2-4, 364:23-365:11.

467. ND_0006830.

468. Hill Report ¶114 (citing Deposition of Elaine Varas (Aug. 2, 2023) 138:3-11).

decisions “to keep Penn aligned with the 568 Presidents’ Group Consensus Methodology.”⁴⁶⁹ Indeed, even close in time to before Penn left the group, its formal financial aid policy manual provided that “Penn uses a variation of the College Board’s Institutional Methodology (IM) to award institutional grant funds. Penn further adheres to the CM, a set of guidelines based on IM issued by the 568 President’s Group of need-blind schools.”⁴⁷⁰ And when deciding to end its formal membership in the 568 Group in 2020, Penn alluded to “the restrictions of the needs analysis consensus document, which requires us to apply certain needs analysis assessment constantly amongst all of our schools.”⁴⁷¹

270. Dr. Hill cites the testimony of Rice’s Former Assistance Vice President and Executive Director of University Financial Aid Services that “We did not follow all of the consensus methodology. I was told by my predecessor that Rice would not support every element of the consensus methodology. And, for example, the home equity was one example that I used.”⁴⁷² But she also testified that there were “not huge variances, but some minor ones” between Rice’s policies and the CM.⁴⁷³ Even on home equity, Rice was following the CM before she left the university.⁴⁷⁴

271. Dr. Hill cites the testimony of Vanderbilt’s Executive Director of Financial Aid that “with regard to, you know, the professional judgment section, you know, and elements of the consensus methodology, we would develop our own procedures that we ultimately followed.”⁴⁷⁵ But contemporaneous qualitative evidence shows that Vanderbilt understood that “The IM formula components that we use is usually referred to as Consensus Methodology (based largely on the work of the 568 group).”⁴⁷⁶ And after signing the 568 Group’s memorandum of understanding, the

469. PENN568-LIT-00069194.

470. PENN568-LIT-00018593 at -632.

471. PENN568-LIT-00137113 at -117.

472. Hill Report ¶114 (citing Deposition of Anne Walker (Sept. 8, 2023) [hereafter Walker Dep.] 147:12-17).

473. Walker Dep. 204:7-16.

474. *Id.* 37:5-24.

475. Hill Report ¶114 (citing Deposition of Brent Tener (July 12, 2023) 76:9-23).

476. Vanderbilt-00204227.

Executive Director of Financial Aid stated that Vanderbilt agreed to “adhere by a set of [e]numerated guidelines” and that “we are in full compliance.”⁴⁷⁷ Vanderbilt’s Vice Provost for Enrollment and Dean of Admissions explained in a 2014 email to the Chancellor that “Vanderbilt benefits” from its participation in the 568 Group, because: “This discussion allows for a common approach, so a family’s expected family contribution does not vary to any great extent from school to school” and thereby “helps to avoid bidding wars between schools.”⁴⁷⁸

272. Dr. Hill cites the testimony of Yale’s former Director of Undergraduate Financial Aid that the 568 Group “was not about doing it the exact same way. It’s about learning how and agreeing on factors in the need analysis that can best support students.”⁴⁷⁹ But Yale specifically left the 568 Group in order to be “free to give families more aid than they would have gotten under the consensus methodology” since the 568 Group imposed “one needs-analysis formula that everyone has to sign on to.”⁴⁸⁰ In addition, “agreeing on factors” is part of the Overarching Agreement.

273. Dr. Hill’s citation of deposition testimony regarding EFC calculations notably omits Georgetown. Georgetown President DeGioia was Chair of the 568 Group from 2009 to the Group’s dissolution in 2022, and Georgetown’s Dean of Student Affairs McWade was Chair of the 568 Group’s Technical Committee for 4 years. DeGioia and McWade have both recognized that the CM is the “common formula” that Georgetown used to calculate EFC.⁴⁸¹ McWade testified that the 568

477. Vanderbilt-00041936.

478. Vanderbilt-00389088.

479. Hill Report ¶114 (citing Deposition of Scott Wallace-Juedes (July 7, 2023) 152:21–153:17).

480. Caitlin Roman, *University Leaves Financial Aid Group*, YALE DAILY NEWS (Sept. 26, 2008), <https://yaledailynews.com/blog/2008/09/26/University-leaves-financial-aid-group>.

481. Deposition of John DeGioia (Feb. 16, 2024) Ex. 13; McWade Dep. at 105:18-19. *See also* Georgetown’s Response Nos. 2-3 to Plaintiffs’ First Set of Requests for Admission (Nov. 13, 2023) (admitting that Georgetown “has used ‘the Presidents’ 568 Working Group’ Consensus Methodology, ‘with modifications, to determine financial need for . . . institutional scholarship funds) (quoting from Georgetown’s Office of Student Financial Services Policies and Procedures Manual). Georgetown adds that its use of the CM is subject to professional judgment, but the 568 Group’s sixth Core Principle makes clear that such judgment applies only to “unique or extenuating circumstances in individual cases . . . [and] is not the proper mechanism for systematically treating . . . students differently in order to advance institutional objectives.” *See* McWade Dep. 75:6-21 and Ex. 3 (discussing 568 Survey showing that 100% of members agree with this Principle).

Group's purpose following the passage of the 568 statute was "to see if we could agree on how to analyze families' ability to pay," and that in turn led to the concern that "after the exemption expired . . . we're going back to the old Wild Wild West where schools would do . . . what they wanted and different things."⁴⁸² McWade also stated in writing: "it was made clear to those wanting to participate in 568 that they could not use an EFC that was lower than" an EFC calculated using the 568 Group's Consensus Approach.⁴⁸³

274. As Dr. Yermack pointed out, Georgetown had the weakest unrestricted endowment.⁴⁸⁴ This is a likely reason why Georgetown wanted the 568 Group to exist, by limiting competition to bring overall spending on financial aid down, thereby bringing other 568 Group members down to the weakest link. In Georgetown's own words: "The participating institutions believe that the Consensus Approach, when applied in a consistent manner, serves to diminish or eliminate the divergent results that threaten the long-standing tradition of awarding aid on the basis of need. This methodology theoretically enables Georgetown to offer aid packages that are more consistent with the other universities who have signed on, and may in part contribute to increased yield rates."⁴⁸⁵

B. Dr. Hill's Qualitative Evidence Regarding Net Price Calculations (Section 5.2.4)

275. Dr. Hill cites certain documents he contends are consistent with Defendants offering different net prices as a result of competing on the packaging element of financial aid to facilitate enrollment. Yet his analysis does not attempt to rebut the qualitative evidence showing that Defendants applied the affordability principle to packaging as well.⁴⁸⁶ In addition, even if one were to accept as a premise that the 568 Group competed on financial aid packages, that limited

482. McWade Dep. 190:18-191:9.

483. GTWNU_0000193182 (McWade Dep. Ex. 11).

484. Yermack Report, Appendix B Table 1.

485. GTWNU_19960 at -063 (2006-07 Report). *See also* GTWNU_0000079934 at -947 (2020 "Financial Aid at Georgetown" presentation repeating statement that the CA "serves to diminish or eliminate the divergent results that threaten the long-standing tradition of awarding aid on the basis of need").

486. *See, e.g.*, Singer Report ¶¶188-191.

competition would not create adequate competition to prevent artificial inflation in the Effective Institutional Price caused by the Challenged Conduct.

276. For instance, Dr. Hill cites a 2015 Briefing Book from Georgetown’s Office of Student Financial Services responding to Georgetown “losing students . . . to those schools offering reduced or no loans.”⁴⁸⁷ But Georgetown understood that this packaging step occurred *after* “we agreed to use the same need analysis methodology as the other” 568 Group members.⁴⁸⁸ Georgetown’s internal conclusions it had “financial constraints” that made it not “offer[] aid packages that are as attractive as our competitors” is still consistent with an agreement to use the underlying methodology.⁴⁸⁹

277. Dr. Hill similarly cites a 2013 Notre Dame email chain where its Director of Admissions discusses setting enrollment goals while acknowledging there were competitors “offering no-loan packages.”⁴⁹⁰ But just two years later, Notre Dame’s Dean of Financial Aid celebrated the renewal of the Section 568 exemption, noting that “this is great news for Notre Dame and how we compete and try to ‘level’ the playing field at least in the calculation of need” even while acknowledging there would still see “packaging differences” with its competitors.⁴⁹¹

278. Dr. Hill also cites a 2013 document from Penn that shows its tracked competitors’ financial aid policies, comparison of net prices, and Penn’s win-loss percentages to those schools.⁴⁹² This reflects that Defendants were competitors. The qualitative evidence also reveals that Penn would later go on to resign from the 568 Group to provide “increased flexibility in our needs analysis”⁴⁹³—flexibility that the document confirms could not be obtained from “competition” in packaging.

487. Hill Report ¶132 (citing GTWNU_0000107168 at -219).

488. GTWNU 0000071253.

489. *Id.*

490. Hill Report ¶132 (citing ND_0158341 at -343).

491. ND_0065243.

492. PENN568-LIT-00167414.

493. PENN568-LIT-00000002.

Similarly, Dr. Hill cites a 2014 document prepared by Emory's Dean of Admission describing that the "more we offer in financial aid . . . the higher our yield will be."⁴⁹⁴ By 2014 Emory had withdrawn from the 568 Group and its strictures.⁴⁹⁵

494. Hill Report ¶132 (citing Emory_568Lit_0058886).

495. Hill Report ¶75.

APPENDIX 3: DEFENDANTS' EXPERTS' PREPACKAGED ANALYSES ARE NOT PERSUASIVE

279. The Defendant Experts present a number of original analyses that do not respond to my Initial Report, but rather attempt to present a new story. I refer to these analyses as “prepackaged” analyses, because they are unresponsive analyses that did not require seeing my Initial Report to create. I present a table detailing these prepackaged analyses. I provide the author of the analysis, the overall finding summary, and note if the finding is noteworthy or not. I also describe why the analysis is not compelling. For those results that I believe are worth noting, I go into further detail below. The Appendix 3 section that responds to the listed prepackaged analysis is also included in Table 8 below.

TABLE 8: SUMMARY OF DEFENDANTS' PREPACKAGED ANALYSES

Author	Finding Summary	Noteworthy Finding?	Rejoinder (with Appendix 3 section number)
Dr. Hill	Difference-in-differences model shows that the amount of institutional grant aid awarded doesn't differ between Defendants and Peer schools. ⁴⁹⁶	Yes	Section B.1
Dr. Hill	Average total real institutional grant aid per Defendant increases over time. ⁴⁹⁷	No	Misconstrues the Counterfactual (A.3)
Dr. Hill	The percentage of full-time, first-time, degree-seeking undergraduates received institutional grant aid increases over time. ⁴⁹⁸	No	Misconstrues the Counterfactual (A.3)
Dr. Hill	The number of Penn students receiving loans decreases over time. ⁴⁹⁹	No	Misconstrues the Counterfactual (A.3)
Dr. Hill	The number of full ride financial aid packages is increasing over time. ⁵⁰⁰	No	Misconstrues the Counterfactual (A.3)
Dr. Stiroh	Cross-admitted students' EFCs vary by school (shown using scatterplots and a table showing the percentage/amount difference). ⁵⁰¹	Yes	Section B.2
Dr. Hill	An example student admitted to four Defendant institutions has four different EFCs. ⁵⁰²	No	Elevates Anecdotal Information over Statistical Analysis (A.1)
Dr. Hill	EFCs vary by more than 20 percent for 56 percent of cross-admitted students. ⁵⁰³	Yes	Section B.3
Dr. Hill	The average EFC spread is between \$15,400 and \$26,600 (depending on the number of Defendants the student is admitted to) for cross-admitted students. ⁵⁰⁴	Yes	Section B.3
Dr. Hill	Average variation of cross-admitted students' EFCs does not significantly change when the	No	Uses a Faulty Benchmark (A.2)

496. Hill Report ¶91, Figure 12.

497. *Id.* ¶¶80-81, Figures 7 and 8.498. *Id.* ¶82, Figure 9.499. *Id.* ¶84, Figure 10.500. *Id.* ¶85, Figure 11.

501. Stiroh Report ¶155, Figures 6.1, 6.2, 6.3, and 6.4.

502. Hill Report ¶99, Figure 13.

503. *Id.* ¶101, Figure 14.504. *Id.* ¶102, Figure 15.

-181-

	Defendant joins or exits the 568 Group. ⁵⁰⁵		
Dr. Hill	Average absolute difference in EFCs for cross-admitted students doesn't significantly change over time. ⁵⁰⁶	No	Uses a Faulty Benchmark (A.2)
Dr. Hill	An example student admitted to four Defendant institutions has four different net prices. ⁵⁰⁷	No	Elevates Anecdotal Information over Statistical Analysis (A.1)
Dr. Hill	Net prices differ by more than 20 percent for 64 percent of cross-admits. ⁵⁰⁸	Yes	Section B.4
Dr. Hill	The average net price for cross-admitted students is between \$19,500 to \$45,500, depending on the number of Defendants who admitted the student. ⁵⁰⁹	Yes	Section B.4
Dr. Hill	Defendants offer students different average net prices compared to the net prices at other Defendants, for students with the same EFC. ⁵¹⁰	Yes	Section C.1, C.2
Dr. Hill	Average variation of cross-admitted students' net prices does not significantly change when the Defendant joins or exits the 568 Group. ⁵¹¹	No	Uses a Faulty Benchmark (A.2)
Dr. Hill	Defendants did not change the total amount of institutional grant aid awarded in the three years upon entering or exiting the 568 Group, compared to non-Defendants. ⁵¹²	No	Uses a Faulty Benchmark (A.2)

505. *Id.* ¶¶104-105, Figure 16. *Id.* ¶111, Figure 18.

506. *Id.* ¶106, Figure 17.

507. *Id.* ¶119, Figure 19.

508. *Id.* ¶120, Figure 20 (showing that net prices are not exact matches for any cross-admitted students, and that net prices are different by more than five percent for 86 percent of cross-admits, and by more than 20 percent for 64 percent of cross-admits.).

509. *Id.* ¶121, Figure 21.

510. *Id.* ¶124, Figure 22.

511. *Id.* ¶126, Figure 23. *Id.* ¶130, Figure 25.

512. *Id.* ¶138, Figure 26. *See also Id.* at n. 200. *Id.* ¶143, Figure 27. *See also Id.* at n. 205.

A. Irrelevant Prepackaged Analyses**1. Dr. Hill Presents Isolated Examples Under the Guise of Analysis**

280. Dr. Hill presents a bar graph showing the EFC for a single student, admitted to four Defendant institutions.⁵¹³ He repeats this with a different student with four different net prices.⁵¹⁴ These are essentially anecdotal information that Dr. Hill uses to introduce more fulsome reviews of the data. These one-off examples present no analysis and are therefore not worth addressing. I review Dr. Hill's claim that overall students have deviations in EFC and net price below.⁵¹⁵

2. Dr. Hill Uses Faulty Benchmarks in His Prepackaged Analyses

281. Dr. Hill conducts several prepackaged analyses that use a faulty benchmark. These include his examination of how exiting or entering the 568 Group impacts total institutional grant aid, EFCs, and net prices.⁵¹⁶ The premise of these analyses is to look at how variation in institutional grant aid, EFCs, and net prices changes the same in the three years when the Defendant is in the 568 Group compared to a three-year period when the Defendant is out of the 568 Group. Dr. Hill claims that the variation does not increase when the Defendant exits the 568 Group (or decrease when the Defendant joins the 568 Group), and that this indicates that Challenged Conduct did not have an effect.

282. Dr. Hill uses the wrong benchmark with these analyses. It is not relevant to assess if the EFC varied more over time. It matters if the common factors agreed upon by the 568 Group members, the Challenged Conduct, led to increased EFCs and corresponding decreased institutional grant aid and increased Effective Institutional Prices. Without a regression analysis, Dr. Hill is once more viewing the (cross-admit subsection of) data without controlling for any other factors. I show

513. Hill Report ¶99, Figure 13.

514. *Id.* ¶119, Figure 19.

515. *See* Appendix 3.B.3, *infra*.

516. *Id.* ¶¶104-105, Figure 16. *Id.* ¶111, Figure 18. *Id.* ¶126, Figure 26. *Id.* ¶130, Figure 25. *Id.* ¶127, Figure 24.

that the overall EFC values indeed were inflated by the Challenged Conduct with my EFC regression analysis in Table 1. I also showed that the Effective Institutional Prices were increased by the Challenged Conduct with my regression analysis in my Initial Report.⁵¹⁷

3. Dr. Hill Repeatedly Misconstrues the Relevant Counterfactual

283. Dr. Hill claims that institutional financial aid increased over time, whereas the alleged collusion between 568 Group members should result in the stagnation or decline in the amount of institutional financial aid awarded over time.⁵¹⁸ Dr. Hill asserts that the increase in institutional aid can be demonstrated by multiple metrics, including the average total real institutional grant aid awarded per Defendant,⁵¹⁹ the percentage of full-time, first-time, degree-seeking undergraduates who received institutional grant aid,⁵²⁰ the number of Penn students receiving loans decreasing,⁵²¹ and the number of “full ride” financial aid packages increasing.⁵²² Dr. Hill also cites documents and testimony from Defendants as support that the amount of institutional financial aid spending has increased over time.⁵²³ Dr. Hill ignores the relevant counterfactual in these assessments.

284. The relevant question is not whether institutional grant aid increased over time, or even during the Class Period. Rather, the question is if, absent the Challenged Conduct, the institutional grant aid would have been *even higher* than the observed level. Increasing institutional grant aid over the Class Period does not mean that there was no conspiracy; nor does it mean that the alleged conspiracy was ineffective. Dr. Hill avoids this relevant metric by only showing that the observed level of institutional grant aid increased over time.

517. Singer Report Table 11.

518. Hill Report ¶86.

519. *Id.* ¶¶80-81, Figures 7 and 8.

520. *Id.* ¶82, Figure 9.

521. Hill Report ¶84, Figure 10.

522. *Id.* ¶85, Figure 11.

523. *Id.* ¶146.

285. Studying simplistic trends does not answer the question of interest, because trends do not isolate the effect of the Challenged Conduct, controlling for other factors that affect Effective Institutional Price. Dr. Hill focuses on how students have fared relative to their predecessors. An increase in institutional aid spending does not indicate the absence of a conspiracy, because the spending could have been even higher in the but-for world. To answer the true question of interest, one needs to conduct a regression, examining how the Effective Institutional Price changes during the Class Period, while holding relevant control factors constant. This is the point of my regression analysis in my Initial Report.⁵²⁴ Dr. Hill fails to consider the corresponding increase in the cost of attendance (by using the Effective Institutional Price). Furthermore, Dr. Hill's review of students receiving full rides is confusing, given that these students would not be Class Members. Finally, Class Members who attended schools that adopted no-loan policies were still injured by the suppression of institutional grant aid made possible by the Challenged Conduct.

B. Relevant Prepackaged Analyses

1. Dr. Hill's Assertion That Defendants Increased Financial Aid at the Same Pace as Peer Schools Is Misleading

286. In Section 4.2 of his rebuttal report, Dr. Hill argues that the "Defendants increased financial aid at the same pace as peer schools after the introduction of the consensus methodology guidelines."⁵²⁵ To support his argument, Dr. Hill compares data for Defendants to data for a group of five non-Defendant schools in the same relevant market as Defendants. The group of non-defendant schools consists of Carnegie-Mellon, Harvard, Princeton, Stanford, and Washington University of St. Louis.

524. See Singer Report Table 11.

525. Hill Report §4.2.

287. Dr. Hill relies on IPEDS data to perform his analysis. As I described in my Initial Report, these data consist of aggregated annual information that each institution of higher education reports to the Department of Education (DoE). In contrast to the structured data that Defendants produced in this litigation, IPEDS does not contain data by student. Each observation consists of an annual summary for that specific institution—for example, total institutional grant aid for 2017 that Yale reported to the DoE. In contrast to Dr. Hill, I relied on the student-level structured data. As a general matter, researchers rely on IPEDS data when they cannot obtain such granular information as produced in this litigation or wish to perform analyses that do not require student-level data.

288. Dr. Hill uses a difference-in-differences (DiD) approach to perform his IPEDS-based analysis. DiD is a quasi-experimental technique that attempts to replicate in an observational setting the conditions that occur in a randomized controlled trial. Defendants' entry into the 568 Cartel and engagement in the Challenged Conduct reflects such an observational setting, where researchers observe policies and their effects, and then seek to uncover any causal relationship between the two, while controlling for factors that may confound such an inference. In contrast, in a RCT setting, researchers randomize treatment assignment: whether one participant receives the treatment or the placebo (control) occurs randomly. Because RCTs seldom apply outside a carefully constructed experimental setting and economists nearly always deal with observational data, DiD relaxes the conditions required for the former to apply in the latter, hence the term "quasi-experimental."

289. Specifically, researchers use DiD in observational settings where the exchangeability condition does not hold: in other words, we cannot "exchange" membership in the treated group (Defendants) and the control group (non-Defendants) as we can in an experimental setting where randomization determines the assignment. In an experimental setting, researchers randomly assign a unit to either the treatment or control group before the experiment occurs. As a result, we would

expect that the averages from the two groups should be identical absent the treatment. In an observational setting, the assignment occurred well in the past. As researchers, we can only look back in time and inquire about causal relationships that generated the observed conditions.

290. Dr. Hill's application of DiD uses non-Defendants as a control for Defendants and attempts to recover the causal effect of the Challenged Conduct on Defendants' financial aid. Such a causal inference relies on a critical assumption—namely, that whatever differences existed between the treatment group (Defendants) and the control group (non-Defendants) would have remained the same during the relevant period had Defendants not engaged in the Challenged Conduct. Known as “parallel trends,” this assumption relaxes the exchangeability assumption in the experimental setting. While we cannot assume that the treatment and control group averages would have been identical absent membership in the Challenged Conduct, if parallel trends hold, DiD assumes that whatever differences exist between the two groups would have also remained absent the Challenged Conduct.

291. As a preliminary point, despite the aggregate nature of such data, Dr. Hill finds that his methodology can inform the effects of the Challenged Conduct on a classwide basis. He concludes “Using IPEDS data and a staggered difference-in-differences model, the amount of institutional grant aid awarded does not significantly differ between Defendants and peer schools (Harvard, Stanford, Princeton, Carnegie Mellon, and WashU).⁵²⁶ He implicitly concurs that common methods (in his case, DiD) and analysis can recover the Challenged Conduct's effect on members of the Class. On this point, it appears that both Dr. Hill and I agree with each other, and we both disagree with Dr. Stiroh.

292. Dr. Hill implements DiD using staggered treatment, because some Defendants received the treatment (joined the 568 Group) in different years. Dr. Hill argues that, “This approach

526. *Id.* ¶91, Figure 12.

[DiD with staggered treatment] has some limitations, but with staggered entries and exits, it is important to use it because the standard difference-in-differences approach is flawed.”⁵²⁷ He describes the following details about his DiD regression:

In order to implement the methodology of Callaway and Sant’Anna (2001) I need to exclude any early exit periods from the control groups, imposing the additional assumption that there is no selection into these exits. This approach uses the “not-yet-treated” as a baseline for comparison, i.e., the control group is composed of schools who, until a given academic-year, had never been part of the alleged conduct. I, therefore, drop the early exit periods from this analysis. Additionally, this analysis bases pre-conduct comparisons on the single year prior to conduct in contrast to the single dummy variable model, which uses all available pre-conduct years.⁵²⁸

While Dr. Hill relegates such critical details to a footnote, they highlight the shortcomings of his approach. As he acknowledges, he limits the pre-conduct comparison period to a single year, substantially and improperly reducing the scope of the data. Finding large standard errors, as he does, is unsurprising given the restrictions he imposes. Basing the pre-conduct benchmark on only one year discards the majority of information and subsequently asks too much of the data. As Dr. Hill himself notes, “this approach has some limitations.”⁵²⁹ His acknowledgment understates the problems with his analysis. Fundamentally, instead of fitting the analysis to the data, he selectively excludes critical data to fit the analysis he wants to perform.

293. Further, Dr. Hill’s staggered DiD approach elides important characteristics of these data and misinterprets the data issue at hand. *First*, he claims that the “standard difference-in-differences approach is flawed.”⁵³⁰ The extent to which standard DiD complicates the identification of the parameter of interest (the conduct coefficient) in the presence of staggered treatment depends on the nature of the treatment. Only in the presence of heterogenous treatment effects does this issue arise. As explained in my Initial Report, the nature of the alleged agreement motivates the conclusion

527. *Id.* ¶89.

528. *Id.* at n. 120.

529. *Id.* ¶89.

530. *Id.* ¶89.

that the conduct represents a homogeneous treatment. After all, the very purpose of entering into such an agreement rests on the use of a common financial need determination policy, which Defendants dubbed “the Consensus Methodology.”

294. *Second*, Dr. Hill’s explanation of the “flaw” with using a single conduct dummy (which I used) underscores my observation that doing so yields a conservative estimate of the impact of the Challenged Conduct. Dr. Hill argues:

A single conduct dummy variable can result in estimated effects with limited interpretability. In particular, schools’ conduct period academic years can be the basis of comparison for other conduct period school-academic years. This means a school could counterfactually increase their grant aid during the conduct period, and *the estimated effect on grant aid could decrease*.⁵³¹

Dr. Hill’s theoretical outcome, as described above, would result in an understated effect. As I interpret his example, a previous member of the alleged conspiracy that subsequently left would act as a control for a current member of the conspiracy. Under this scenario, the effect of the conspiracy on a current member’s institutional aid offering would be compared to the institutional aid offered by a previous member of the conspiracy that had subsequently left. To the extent that any residual effects of the conspiracy would remain, the “control” would show restricted levels of institutional aid as a result of such lingering effects. As such, my estimates would understate the causal effect of the conspiracy.

295. *Third*, Dr. Hill offers no analysis of parallel trends. As Sun and Abraham explain (a paper to which Dr. Hill cites as reference for DiD):

If an application includes never-treated units...we need to especially consider whether these never-treated units satisfy the parallel trends assumption. Never-treated units are likely to differ from ever-treated units in many ways, and may not share the same evolution of baseline outcomes. If the never-treated units are unlikely to satisfy the parallel trends assumption, then we should exclude them from the estimation to avoid violation of this assumption.⁵³²

531. *Id.* n. 118.

532. Liyang Sun and Sarah Abraham, *Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects*, 225(2) JOURNAL OF ECONOMETRICS 175-199, 178 (2021).

Dr. Hill's analysis includes non-Defendants, which are by definition non-treated units as they never partook in the alleged conspiracy. Yet, Dr. Hill never investigates the extent to which parallel trends apply. Indeed, the term "parallel" never appears in his report. That non-Defendants may occupy the same relevant market does not imply parallel trends, as Dr. Hill assumes without evidence.

296. *Fourth*, both Borusiyak & Kirill and Sun & Abraham discuss specific staggered designs in which the treatment is an "absorbing state."⁵³³ An absorbing state occurs when the treatment condition occurs as a series of zeroes (no treatment) followed by a series of ones (treatment) but state never returns back to zero. In other words, once treated, the treatment status cannot revert to untreated. Absorbing state treatment has wide applicability: consider the effect of a drug on health care outcomes or of the policy such as a road expansion on traffic congestion. Once treatment has occurred, reversion to non-treated status is unlikely or impossible.

297. Treatment in this particular case is not an absorbing state. Defendants could enter, leave, and reenter, as Yale did. In this case, the sequence would reflect a series of zeroes, followed by a series of ones, followed by a series of zeroes, and then again by a series of ones. As such, Dr. Hill's references do not support his application of DiD in this case. Accordingly, his DiD regression is fundamentally unreliable because it (1) ignores the granular student-level structured data, (2) excludes years of data from IPEDS, (3) misapplies the literature, and (4) fails to perform the requisite checks to investigate whether his analysis does not violate the basic assumptions of DiD (parallel trends). Most importantly, as Dr. Hill appears to acknowledge, the presence of such staggered treatment would result in my regression model underestimating the true effect of the Challenged Conduct.

533. Hill Report ¶89.

298. Finally, as Baker et al. note, researchers should conduct staggered DiD and interpret its results with caution:

The prevalent use of staggered DiD reflects a common belief among researchers that such designs are more robust, and mitigate concerns that contemporaneous trends could confound the treatment effect of interest. However, recent advances in econometric suggest that standard DiD regression estimates with staggered treatment timing often do not provide valid estimates of the causal estimands of interest to researchers—such as the average treatment effect on the treated (ATT)—even under random assignment of treatment... recent work in econometric theory casts doubt on the validity of the TWFE DiD estimator when it is applied to settings with variations in treatment timing. Significant biases may arise when such staggered DiD estimators are used for producing static or dynamic treatment effect estimates.⁵³⁴

299. I do not intend my critique of Dr. Hill's analysis in this section as an indictment of the use of staggered DiD. Researchers commonly apply this methodology under the appropriate conditions, as described above. The instant case does not reflect such conditions, however, rendering Dr. Hill's use of staggered DiD inappropriate. As such, his attempt to use staggered DiD to conclude that he does not find evidence of an effect on net price is inapposite. Though his staggered DiD analysis is misplaced, it serves to support the implication that he agrees that using a common, formulaic methodology can inform classwide effects.

2. Dr. Stiroh's Review of Cross-Admitted Students Is Without Merit

300. Dr. Stiroh argues that individual students admitted to more than one Defendant university (cross-admitted students) often received different EFC calculations from each school to which they were admitted.⁵³⁵ She provides simplistic scatterplots showing the EFCs for cross-admitted students to two given schools and points out that there is variation between the EFCs calculated by each Defendant.⁵³⁶ Dr. Stiroh then provides a table showing the percentage of pairwise combinations with EFC exactly matching, within a ten percent difference, within more than a ten

534. Andrew C. Baker, David F. Larcker, and Charles C.Y. Wang, *How Much Should We Trust Staggered Difference-In-Differences Estimates*, 144(2) JOURNAL OF FINANCIAL ECONOMICS 370-395, 371, 373 (2022).

535. Stiroh Report ¶153.

536. *Id.* ¶155, Figures 6.1, 6.2, and 6.3.

percent difference, within a \$1,000 difference, and more than \$1,000 difference. On average, 74 percent of these pairwise combinations show up with more than a ten percent difference.⁵³⁷ Dr. Stiroh does not provide an explanation as to why she chose the thresholds of ten percent or \$1,000, nor does she bother to test the statistical significance of these results.

301. These calculations are not a proper analysis. Dr. Stiroh is essentially claiming that the EFCs would be expected to match *exactly* for cross-admitted students at various Defendant universities. Despite a reasonable amount of grouping near the 45-degree line, particularly for low EFC values (in the bottom left corner of the figures), Dr. Stiroh concludes the data shows that “EFCs calculated by different Defendant schools differ substantially.”⁵³⁸ Eyeballing data in this manner is clearly an insufficient method.

302. Furthermore, there will always be a degree of variation with data such as these. The issue at hand is not whether Defendants agreed to offer identical awards by calculating identical EFCs, but rather whether the Overarching Agreement restricted competition. To address that issue, one must consider variation outside the Conduct Period compared to the variation during the Conduct Period. In Table 1 above, I perform an EFC regression analysis, showing that common factors explain the vast amount of the variation in EFCs, and that the Challenged Conduct was associated with a statistically and economically significant increase in EFC. As additional support, my cross-admit EFC regression in Table 2 above indicates that the EFCs offered to cross-admitted students are closer together when the Defendants are both in the 568 Group.

3. Dr. Hill’s Cross-Admit EFC Analysis Does Not Address The Challenged Conduct

303. Dr. Hill expounds on the fact that the EFCs of cross-admitted students do not match across Defendant institutions. He first claims that the EFCs vary by more than 20 percent for 56

537. *Id.* ¶157, Figure 6.4.

538. *Id.* ¶156.

percent of cross-admitted students.⁵³⁹ He then argues that the average EFC spread among cross-admitted students is between \$15,400 and \$26,600 depending on the number of Defendant schools which admitted the student.⁵⁴⁰ Dr. Hill discusses these points as if he expected that the EFCs should match perfectly for all students across all Defendants. Variations in EFCs are to be expected, however, and do not provide evidence inconsistent with the alleged conspiracy.

304. Variation of EFCs across Defendants, even for the same student, is to be expected. There is no requirement for the EFCs to be the same across all cross-admitted students for the alleged conspiracy to be effective. What matters is whether the Challenged Conduct restricted competition such that it reduced the variance in EFCs for cross-admitted students and whether there is a common set of factors, which translates into harm across all or almost all Class Members. I demonstrate this impact with my cross-admit regression analysis in TABLE 2 and TABLE 3 above, by showing that the EFCs and the resulting Effective Institutional Prices are closer together when Defendants are in the 568 Group compared to when they are not. Furthermore, my EFC regressions in Table 1 show that the Challenged Conduct is associated with a statistically significant overall increase in EFC values.

305. Additionally, Dr. Hill's test of the EFC spread for cross-admitted students emphasizes the percentage of students whose EFCs vary. He does not show, however, the distribution of these variations. Looking at the absolute value, just under 30 percent of cross-admitted students had EFCs that were within \$2,500, 69.1 percent had EFCs within \$10,000, and 91.9 percent had EFCs within \$30,000 of each other.⁵⁴¹

539. Hill Report ¶101, Figure 14.

540. *Id.* ¶102, Figure 15.

541. See my workpapers for details.

4. Dr. Hill Repeats His Inadequate Cross-Admit Analysis With Net Price

306. Dr. Hill rehashes the same arguments for net price. In particular, he argues that cross-admitted students have variation in their net prices at Defendant schools,⁵⁴² and claims that the average net price for cross-admitted students varies by \$19,500 to \$45,500 based on how many Defendants admitted the student.⁵⁴³ Finally, Dr. Hill claims that Defendants offer students different average net prices from those offered at other Defendants to students with the same EFC. He concludes that differences in EFC and packaging must be causing different net prices across Defendants.⁵⁴⁴

307. As I have pointed out several times, net prices (like EFC) are not expected to match exactly between Defendants. There is no requirement for the net prices to be the same across all cross-admitted students for economics to suggest that a conspiracy exists. My Effective Institutional Price cross-admit analysis, in TABLE 3 above, shows that the Effective Institutional Prices have less variation when Defendants are in the 568 Group. My cross-admit analysis specifically controls for the number of Defendants that admitted the student, indicating that regardless of the number of Defendants admitting a student, the Effective Institutional Prices are closer together when the Challenged Conduct is in effect.

308. To Dr. Hill's argument that students with the same EFCs are offered different net prices by different Defendants, it is important to first note that the Challenged Conduct operates with a variety of mechanisms and does not rely entirely on EFC translating into matching net prices. The final Effective Institutional Price is what matters to students, so if EFC doesn't translate perfectly

542. Hill Report ¶120, Figure 20 (showing that net prices are not exact matches for any cross-admitted students, and that net prices are different by more than five percent for 86 percent of cross-admits, and by more than 20 percent for 64 percent of cross-admits.).

543. *Id.* ¶121, Figure 21.

544. *Id.* ¶124, Figure 22.

into price, my Effective Institutional Price regressions in my Initial Report still show that the Challenged Conduct is associated with higher prices being charged to students.⁵⁴⁵ Furthermore, my cross-admit regressions in TABLE 2 and TABLE 3 above show that the Challenged Conduct is associated with a statistically significant narrower range in EFCs and Effective Institutional Prices. The cross-admit regression results indicate that even if students are not assigned identical EFCs by different Defendants, the Challenged Conduct is associated with EFCs that are closer together.

C. Defendants' Experts' Claims That The Consensus Methodology Did Not Standardize Financial Aid Are Without Merit

1. Dr. Hill's Argument That Defendants Do Not Have Matching Net Prices Is Irrelevant and Misleading

309. Dr. Hill claims that there is no significant variation in net price when Defendants join or exit the 568 Group when examining Defendants who entered or exited with three years of data both in and out of the 568 Group.⁵⁴⁶ He additionally claims that the average variation of net prices offered to cross-admitted students increases after the introduction of the CM.⁵⁴⁷

310. Dr. Hill unnecessarily limits his data to the three years on either side of a Defendant entering or exiting the 568 Group, rather than evaluating every year that exists in each Defendant's data production. He also fails to control for all of the other factors that could affect the net price. This analysis is simplistic and is cherry-picking of the highest order. It is vastly inferior to my regression approach, presented in Table 11 of my Initial Report, which included all available data and controlled for potentially confounding factors.⁵⁴⁸

545. Singer Report Table 11.

546. Hill Report ¶126, Figure 23. Hill Report ¶130, Figure 25.

547. *Id.* ¶127, Figure 24.

548. Singer Report Table 11.

2. Dr. Hill's Claim That Defendants Did Not Change Their Financial Aid When Entering or Exiting the 568 Group Is Incorrect

311. Dr. Hill argues that Defendants did not change the total amount of institutional grant aid awarded in the three years upon entering or exiting the 568 Group, compared to non-Defendants.⁵⁴⁹ There are a number of issues with this analysis. *First*, Dr. Hill considers Defendants that were not in the 568 Group in the year being evaluated as a control group. While I conservatively do not include these Defendants in my Initial Report's Effective Institutional Price regression analysis, they were likely still impacted by the effects of the 568 Group and thus cannot be considered a reliable control group. Showing that these Defendants awarded similar amounts of aid when not part of the 568 Group to Defendants who had recently exited does not prove that these Defendants are an "appropriate control group," as Dr. Hill claims.⁵⁵⁰ *Second*, he again must limit his analysis to only three years, due to the availability of data. *Third*, he focuses on both the incorrect metric and the irrelevant issue.

312. The relevant question is if the Challenged Conduct resulted in Class Members being charged higher Effective Institutional Prices than they would have been in the absence of the Challenged Conduct. Dr. Hill continuously attempts to reframe this question. By looking at the total aid awarded and controlling for nothing, Dr. Hill merely shows that Defendants awarded similar amounts of aid as each other (and a few additional non-Defendants). Directly answering the relevant question by conducting a regression and controlling for relevant factors is what I did in my Initial Report.

549. Hill Report ¶138, Figure 26. *See also id.* at n. 200. *Id.* ¶143, Figure 27. *See also Id.* at n. 205.

550. *Id.* ¶143.

**APPENDIX 4: STATISTICAL ANALYSES OF THE REMAINING DEFENDANTS' WEALTH FAVORITISM
IN ADMISSIONS**

313. I summarize below my analyses of the remaining (non-settling) Defendants' data regarding admissions of donor-linked applicants.⁵⁵¹ For each Defendant, I examine the acceptance rates for applicants that Counsel has indicated to me appear on certain priority lists for admissions ("priority-designated students" or "priority-designated applicants"), with the understanding that Defendants place certain applicants on such priority lists because they are related to high-value donors or potential donors, among other reasons. I then compare the acceptance rates for the priority-designated applicants against those against the acceptance rate for non-priority-designated students or the general acceptance rate (where the non-priority-designated student acceptance rate is not available) in the same admissions cycle. I also compare the standardized test (SAT and ACT) scores for priority-designated students with those of non-priority-designated students.

314. In sum, for each remaining Defendant, priority-designated students were statistically significantly more likely to be admitted than non-priority-designated students. In addition, for all remaining Defendants, priority-designated students had statistically significantly lower or comparable standardized test scores compared to non-priority-designated students.

A. Cornell

315. Counsel produced a spreadsheet containing two lists. The first contains UIDs that were present on Cornell's admissions watchlist. The second contains the UIDs from Cornell's watchlist ("Watchlist applicants") as well as the UIDs from the Cornell Connection form.⁵⁵²

551. I understand that Cornell, Georgetown, MIT, Notre Dame, and Penn have not settled.

552. There are 16 UIDs that appear only in the first list. Since the second list is intended to include all of the first list, I added these 16 UIDs to the second list before examining the data.

316. Cornell produced admissions data for the academic years 2008-2024.⁵⁵³ The Cornell Connection and/or watchlist UIDs appear in the data between 2008-2023, so 2024 is excluded from the comparison analysis. I use the latest admissions data, as used by Dr. Hill.⁵⁵⁴

317. For the years in which (1) Cornell provided admissions data and (2) there is at least one priority-designated applicant, I provide the following table comparing the admission rate for priority-designated applicants with the admission rate of all other applicants. I separate the priority-designated applicants into two groups: Cornell Watchlist applicants and Cornell Watchlist plus Cornell Connection applicants.⁵⁵⁵

318. The admission rate is defined as the total admitted applicants divided by the total number of applicants. Column 1 indicates the admission rate for Cornell Watchlist applicants alone. Column 2 shows the admission rate for Cornell Watchlist plus Cornell Connection applicants combined. Column 3 shows the admission rate for all other applicants (applicants who are not on the Cornell Watchlist or the Cornell Connection list).

553. The academic year corresponds to the fall through summer semesters defined using the year associated with the fall semester. For instance, the 2018 academic year corresponds to fall 2018 through summer 2019.

554. See my workpapers for details.

555. I understand that Cornell Connection applicants are all legacy applicants, whereas the Watchlist consists of applicants of interest to university personnel and faculty such as the president, deans, and faculty. See Deposition of Jason Locke (Dec. 11, 2023) 178:1-181:12.

APPENDIX 4 TABLE 1: CORNELL ANNUAL ADMISSIONS RATES BY APPLICANT GROUP

	[1]	[2]	[3]
Academic Year	Watchlist Applicants	Watchlist + Connection Applicants	All Other Applicants
2008	100%	100%	32%
2009	100%	50%	22%
2010	80%	80%	19%
2011	100%	100%	18%
2012	67%	67%	17%
2013	100%	100%	16%
2014	100%	38%	14%
2015	82%	71%	15%
2016	87%	57%	15%
2017	79%	40%	14%
2018	67%	51%	10%
2019	78%	42%	11%
2020	50%	31%	11%
2021	64%	65%	9%
2022	67%	66%	7%
2023	81%	81%	8%

319. These findings show that the admission rate of Cornell Watchlist and the admission rate of Cornell Watchlist plus Cornell Connection applicants are higher than the admission rate of all other applicants in every year with available data. After examining these findings, I conducted a t-test to assess whether these differences are significantly different. I found that both Cornell Watchlist and Cornell Watchlist plus Cornell Connection applicants are admitted at a statistically significant higher rate than all other applicants.⁵⁵⁶ Priority-designated applicants had admission rates that were at least 20 percentage points higher than non-priority applicants in all years with available data.⁵⁵⁷

556. I conducted a t-test to compare the admission rates of Cornell Watchlist applicants to all other applicants between 2008 and 2023. Watchlist applicants were admitted at a statistically significant higher admission rate compared to all other applicants, at a one percent significance level. I conducted another t-test to compare the admission rates of Cornell Watchlist and Cornell Connection with all other applicants between 2008 and 2023. Watchlist and Connection applicants were admitted at a statistically significant higher admission rate compared to all other applicants, at a one percent significance level. *See* my workpapers for details.

557. In 2020, 31 percent of Watchlist or Connection applicants were admitted while only 11 percent of all other applicants were admitted. All other years see greater differences in the respective admission rates.

Across all available years, the average difference in admission rates between Cornell Watchlist or Cornell Connection priority-designated applicants and non-priority applicants is 50 percentage points.⁵⁵⁸ In sum, Cornell admitted priority-designated applicants at a significantly higher rate than the rest of its undergraduate applicant pool.

320. I also test whether the priority-designated applicants who were admitted had higher standardized test scores relative to the non-priority-designated students who were admitted. I used Cornell's admissions data and identified composite SAT and ACT scores. Where these scores are available, I calculated the average composite SAT and ACT score per academic year for each group. Appendix 4 Table 2 shows the average SAT scores for the same groups. Appendix 4 Table 3 shows the average ACT scores for admitted watchlist, watchlist or Cornell Connection, and all other students.

558. $((100\% - 32\%) + (50\% - 22\%) + (80\% - 19\%) + (100\% - 18\%) + (67\% - 17\%) + (100\% - 16\%) + (38\% - 14\%) + (71\% - 15\%) + (57\% - 15\%) + (40\% - 14\%) + (51\% - 10\%) + (42\% - 11\%) + (31\% - 11\%) + (65\% - 9\%) + (66\% - 7\%) + (81\% - 8\%)) / 16 = 50$ percent.

APPENDIX 4 TABLE 2: CORNELL AVERAGE ANNUAL SAT SCORES BY ADMITTED GROUP

	[1]	[2]	[3]
Academic Year	Watchlist Students	Watchlist + Connection Students	All Other Students
2017	1,347	1,365	1,393
2018	1,460	1,443***	1,484
2019	1,290	1,426***	1,486
2020	1,150	1,357***	1,480
2021	1,500	1,495	1,489
2022	1,514	1,512	1,506
2023	1,420	1,420	1,516

Notes: *** indicates p-value<0.01, ** indicates p-value<0.05, * indicates p-value<0.1 for the t-test when compared to the SAT score for all other students.

APPENDIX 4 TABLE 3: CORNELL AVERAGE ANNUAL ACT SCORES BY ADMITTED GROUP

	[1]	[2]	[3]
Academic Year	Watchlist Students	Watchlist + Connection Students	All Other Students
2008	30.7		31.5
2009			31.2
2010	32.0		31.5
2011	30.0	30.0	31.5
2012	30.0		31.7
2013			31.8
2014		33.3	32.1
2015	29.8*	29.8**	32.2
2016	33.2	32.4	32.4
2017	31.1*	31.0**	32.6
2018	34.0	32.3*	33.0
2019	33.8	33.2	33.2
2020		30.0**	33.1
2021	33.9	33.9	33.4
2022	33.8	33.8	33.6
2023	33.0	33.0	34.0

Notes: *** indicates p-value<0.01, ** indicates p-value<0.05, * indicates p-value<0.1 for the t-test when compared to the ACT score for all other students.

321. Appendix 4 Table 2 shows that the average ACT score for priority-designated students is approximately equal to that of non-priority-designated students. Of the twelve years with ACT data for priority-designated and non-priority-designated students, six have non-priority-designated students scoring better on average while six years have the priority-designated students scoring better

on average. The only examples where there is a statistically significant score difference between groups is when the ACT score is lower for admitted priority-designated students.⁵⁵⁹

322. Appendix 4 Table 3 shows that the average SAT score is higher for non-priority-designated students in every year except for 2022, when Watchlist plus Connection (or Watchlist alone) students average approximately 8 points better (out of a possible 1600). The only statistically significant differences occur when the admitted priority-designated students score lower than the non-priority-designated students.⁵⁶⁰ They are being admitted in spite of scores that are lower than those of non-priority admitted students.

323. The results combined in Appendix 4 Tables 1-3 establish that, on average, from 2008-2023, Cornell's priority-designated applicants (a) were admitted at higher rates than other, non-priority-designated applicants, and (b) were admitted with statistically similar or lower average standardized test scores than other, non-priority-designated students. I conclude from these data that some factor other than average ACT or SAT score is positively influencing the admission of priority-designated students.

B. Georgetown

324. In my review of Georgetown's structured data, I observed applicant designations, to which Georgetown refers as "recruitment categories."⁵⁶¹ I classify the recruitment categories

559. There is a statistically significant difference between Watchlist and Connection students and all other students in 2015, 2017, 2018, and 2020. These values are significant at the five percent level (except for 2018, which is significant at the ten percent level). For the watchlist students alone, there is a statistically significant difference at the ten percent level in 2015 and 2017. This indicates that the priority-designated students had statistically lower ACT scores during these years. *See* my workpapers for details.

560. There is a statistically significant difference between watchlist and connection students and all other students in 2018, 2019, and 2020. These values are significant at the one percent level. This shows that the priority-designated students had statistically lower SAT scores during these three years. *See* my workpapers for details.

561. *See* Georgetown's August 8, 2023, Structured Data Responses ("2023.08.08 Attachment C_Undergraduate Admissions Codes_Recruitment Categories (CONFIDENTIAL-SUBJECT TO PROTECTIVE ORDER)").

“President’s List (Final),” “President’s List (Tracking),” and “Office of Advancement” as priority-designated applicant groups.⁵⁶²

325. Georgetown provided admissions data containing recruitment categories from 2018 through 2023.⁵⁶³ Appendix 4 Table 4 provides admissions rates for each applicant group by academic year.⁵⁶⁴ The admissions rate is defined as the total admitted applicants divided by the total number of applicants. Columns 1-3 show admissions rates for priority-designated applicant groups. Column 4 shows admissions rates for all other students that are not priority-designated.⁵⁶⁵

APPENDIX 4 TABLE 4: GEORGETOWN ANNUAL ADMISSIONS RATES BY APPLICANT GROUP

	[1]	[2]	[3]	[4]
Academic Year	President's List (Final)	President's List (Tracking)	Office of Advancement	All Other Students
2018	86.7%	37.9%	45.5%	11.9%
2019	83.0%	42.9%	43.3%	10.7%
2020	87.3%	51.9%	51.9%	13.3%
2021	90.2%	49.3%	46.9%	9.3%
2022	100.0%	42.1%	46.0%	10.0%
2023	89.7%	48.2%	50.2%	11.4%

326. The admissions rates for each priority-designated applicant group are much greater than the admissions rates for non-priority-designated applicant groups in all academic years, and these differences are all statistically significant at a one percent significance level.⁵⁶⁶ President’s List

562. These include recruitment category codes “PREG,” “PRSF,” “PRSW,” and “PRST” (which refer to President’s List applicants) and “OAU” (which refers to Office of Advancement applicants).

563. Georgetown’s admissions data do not contain recruitment categories prior to the 2018 academic year.

564. The academic year corresponds to the fall through summer semesters defined using the year associated with the fall semester. For instance, the 2018 academic year corresponds to fall 2018 through summer 2019.

565. Georgetown’s data include ten separate “decision” variables that track multiple decisions made over time for each applicant. I define an applicant as having been “admitted” if any of these decision variables contain the terms “Admit” or “Deposit Paid.” I define an academic year as the fall through summer semesters corresponding to the fall year. For example, the 2019 academic year corresponds to fall 2019–summer 2020.

566. For each priority-designated applicant group, I conducted a t-test comparing their admit rates to the admit rates for all non-priority-designated applicants between 2018 and 2023. I found that priority-designated applicants were admitted at a statistically significantly higher admissions rate compared to non-priority students at a 1 percent significance level. See my workpapers for details.

(Final) admissions rates range from 83 percent to 100 percent, President's List (Tracking) admissions rates range from 38 percent to 52 percent, and Office of Advancement admissions rates range from 43 percent to 52 percent. In contrast, non-priority-designated admissions rates range from 9 percent to 13 percent.

327. I also test whether the priority-designated applicants who were admitted had higher standardized test scores relative to the non-priority-designated applicants who were admitted. To do this, I use Georgetown's test score data, which provide SAT and ACT scores for Georgetown applicants. For each admitted group, I calculate the average SAT and ACT score per academic year. Appendix 4 Table 5 provides average SAT scores. Appendix 4 Table 6 provides average ACT scores.

APPENDIX 4 TABLE 5: GEORGETOWN AVERAGE ANNUAL SAT SCORES BY ADMITTED GROUP

	[1]	[2]	[3]	[4]
Academic Year	President's List (Final)	President's List (Tracking)	Office of Advancement	All Other Students
2019	1,340	1,368	1,411	1,469
2020	1,373	1,394	1,432	1,470
2021	1,387	1,396	1,412	1,483
2022	1,339	1,376	1,423	1,484
2023	1,375	1,397	1,432	1,474

Notes: I exclude 2018 from this analysis because the test score data for 2018 were incomplete. There is only a single student-ACT score observation for President's List (Final) and for President's List (Tracking) in 2018, and no SAT scores for these two admit groups in 2018. There are also no Office of Advancement SAT or ACT scores for 2018.

APPENDIX 4 TABLE 6: GEORGETOWN AVERAGE ANNUAL ACT SCORES BY ADMITTED GROUP

	[1]	[2]	[3]	[4]
Academic Year	President's List (Final)	President's List (Tracking)	Office of Advancement	All Other Students
2019	31.5	31.7	32.6	32.9
2020	30.5	31.6	32.6	33.0
2021	31.6	32.2	32.5	33.4
2022	30.6	31.0	32.2	33.3
2023	30.6	31.8	32.1	33.0

Notes: I exclude 2018 from this analysis because the test score data for 2018 were incomplete. There is only a single student-ACT score observation for President's List (Final) and for President's List (Tracking) in 2018, and no SAT scores for these two admit groups in 2018. There are also no Office of Advancement SAT or ACT scores for 2018.

328. Appendix 4 Table 5 shows that the average SAT score for each priority-designated admit group (columns 1-3) was always less than the average SAT score for other admitted students (column 4) in every academic year that the data were available. Appendix 4 Table 6 shows that the average ACT score for each priority-designated admit group (columns 1-3) was always less than the average ACT score for other admitted students (column 4) in every academic year that the data were available. In fact, each priority-designated admit group had on average lower SAT and ACT scores than other admitted students in every academic year. I find that these differences between priority-designated and non-priority-designated students' SAT and ACT scores are statistically significant at a five percent significance level for all priority-designated admit groups and academic years available,⁵⁶⁷ except for Office of Advancement priority list's ACT scores in 2019 and 2020.⁵⁶⁸ While Office of Advancement ACT scores (in Appendix 4 Table 6 column 3) are lower than non-priority-

567. I conduct t-tests by year comparing average SAT and ACT scores of each priority-designated group of students to non-priority-designated students, where the data was available. I am unable to conduct this test for 2018 SAT and ACT scores given the limited test score data during that year.

568. I conduct these t-tests under the null hypothesis that the priority-designated average score is equal to the non-priority-designated average score. The p-values found when comparing Office of Advancement ACT scores to non-priority-designated ACT scores for 2019 and 2020 are 0.26 and 0.29, respectively. It follows that one cannot establish that Office of Advancement ACT scores were different than non-priority-designated ACT scores in 2019 and 2020 at a statistically significant level (based on a five percent significance level). All other admit groups and academic years produce results showing that priority-designated admits had lower average SAT and ACT scores in each year using at least a five percent significance level. *See* my workpapers for details.

designated ACT scores (in Appendix 4 Table 6 column 4) for 2019 and 2020, this difference is not statistically significant at a five percent significance level, meaning that I cannot rule out the possibility that Office of Advancement and non-priority-designated applicants had approximately equal ACT scores during those years.

329. The results combined in Appendix 4 Tables 4-6 establish that, on average, from 2018-2023, Georgetown priority-designated applicants (a) were admitted at statistically significantly higher rates than other, non-priority-designated applicants, and (b) were admitted with standardized test scores that were either less than or equal to other, non-priority-designated applicants. I conclude from these data that some factor other than average ACT or SAT scores is positively influencing the admission of priority-designated students.

C. MIT

330. Counsel provided me with spreadsheets pertaining to the admission status of certain applicants. They include designations assigned to applicants deemed “senior case,” “senior acknowledgment,” and “dean’s case.” I understand that the applicants on these lists are considered to be “priority” applicants for admissions purposes (“priority-designated applicants”).

331. MIT provided admissions data for academic years 2004 through 2024 (with negligible data for 2002 and 2003).⁵⁶⁹ MIT provided lists of the priority-designated applicants for academic years 2011 through 2023. I therefore use the years 2011 to 2023 for the following analysis. I used the admissions data as cleaned by Dr. Hill in conjunction with the MIT priority lists to determine which students were admitted and in which years.⁵⁷⁰

569. The academic year corresponds to the fall through summer semesters defined using the year associated with the fall semester. For instance, the 2018 academic year corresponds to fall 2018 through summer 2019.

570. See my workpapers for details.

332. For years in which (1) MIT provided admissions data and (2) there is at least one priority-designated applicant, I provide the following table comparing the admission rates for priority-designated applicants and all applicants. In the priority-designating spreadsheets, MIT only provided UIDs to uniquely identify applicants who were ultimately admitted to MIT. This means that I must provide the admission rate for all applicants rather than for all non-priority applicants. If I used the UIDs to identify only the priority-designated applicants who were admitted in the complete structured admissions data, I would be underestimating the true admission rate of all other applicants, because the number of rejected applicants would include priority-designated applicants.

333. The admissions rate is defined as the total admitted applicants divided by the total number of applicants. Column 1 indicates the admissions rate for priority-designated applicants. Column 2 shows the admissions rate for all students.

APPENDIX 4 TABLE 7: MIT ANNUAL ADMISSIONS RATES BY APPLICANT GROUP

	[1]	[2]
Academic Year	Priority Applicants	All Applicants
2011	58.3%	9.7%
2012	59.0%	8.8%
2013	52.1%	8.1%
2014	47.5%	7.3%
2015	41.1%	7.8%
2016	25.0%	7.4%
2017	33.3%	6.7%
2018	17.4%	6.4%
2019	18.3%	6.4%
2020	7.5%	3.4%
2021	11.5%	4.2%
2022	13.5%	3.9%
2023	4.7%	4.7%

334. After examining these findings, I conducted a t-test to compare the admission rate of priority-designated applicants with the admission rate of all applicants. I found that priority-

designated applicants are admitted at a statistically significant higher rate than all applicants.⁵⁷¹ That is, MIT admitted priority-designated applicants at a higher rate than the corresponding admission rate of its entire undergraduate applicant pool. Priority-designated applicants were admitted at up to 50.2 percentage point higher rates than comparable non-priority applicants.⁵⁷² The only year in which priority-designated applicants were not admitted at a higher rate than all applicants was 2023. Across all available years, the average difference in admissions rates between the priority-applicants and non-priority applicants is 23.4 percentage points.⁵⁷³

335. I also test whether the priority-designated applicants who were admitted had higher standardized test scores relative to the non-priority-designated applicants who were admitted. Because this test analyzes only the admitted applicants, I can separate out the priority-designated admits from all other admits (unlike in the admission rate comparison). To do this, I used MIT's available data on the SAT and ACT scores of applicants.⁵⁷⁴ For each admitted group (priority and all other students), I calculate the average composite SAT and ACT score per academic year. Appendix 4 Table 8 provided average SAT scores. Appendix 4 Table 9 provides average ACT scores.

571. I conducted a t-test to compare the admission rates of priority-designated applicants compared to all applicants between 2011 and 2023. I found that priority-designated applicants were admitted at a statistically significant higher admission rate compared to all students, at a 1 percent significance level. *See* my workpapers for details.

572. In 2012, priority-designated applicants had a 59 percent admission rate, while all applicants had an 8.8 percent admission rate.

573. $((58.3 - 9.7) + (59.0 - 8.8) + (52.1 - 8.1) + (47.5 - 7.3) + (41.1 - 7.8) + (25.0 - 7.4) + (33.3 - 6.7) + (17.4 - 6.4) + (18.3 - 6.4) + (7.5 - 3.4) + (11.5 - 4.2) + (13.5 - 3.9) + (4.7 - 4.7)) / 13 = 23.4$ percent.

574. As with the admission-rate comparison, I use Dr. Hill's cleaned MIT admissions data to determine which students were admitted. Since Dr. Hill's data does not include the SAT or ACT scores of students, I separately cleaned those MIT admission data files. *See* my workpapers for details.

APPENDIX 4 TABLE 8: MIT AVERAGE ANNUAL SAT SCORES BY ADMITTED GROUP

	[1]	[2]
Academic Year	Priority Students	All Other Students
2017	1,511	1,504
2018	1,540	1,526
2019*	1,568	1,533
2020	1,580	1,533
2021	1,543	1,534
2022	1,525	1,542
2023	1,540	1,539

Notes: Academic year 2019 has a statistically significant difference between priority students' SAT scores and all other students' SAT scores at the 10% level.

APPENDIX 4 TABLE 9: MIT AVERAGE ANNUAL ACT SCORES BY ADMITTED GROUP

	[1]	[2]
Academic Year	Priority Students	All Other Students
2011	32.8	33.4
2012	33.4	33.5
2013	34.3	33.7
2014	33.0	33.7
2015	33.4	33.8
2016	34.3	34.1
2017*	33.4	34.2
2018	35.0	34.5
2019	34.2	34.8
2020	34.0	35.0
2021	35.0	34.8
2022	34.0	35.0

Notes: Academic year 2017 has a statistically significant difference between priority students' ACT scores and all other students' ACT scores at the 10% level.

336. Appendix 4 Table 8 shows that the average SAT score for the priority-designated admit group is only marginally higher than the SAT score for non-priority students in six of the seven years with available data. The difference in the two groups' average SAT scores is between 17 points lower and 47 points higher (out of 1600 possible points), meaning that in any year with available data, priority-designated students have an average SAT score that is no more than three percent higher

than the average SAT score for non-priority students.⁵⁷⁵ A t-test shows that there is no statistically significant difference between the average SAT scores of priority and non-priority students at the 5 percent level.⁵⁷⁶

337. Appendix 4 Table 9 shows that, over the twelve years with available ACT data, the priority-designated admits have lower ACT scores on average than non-priority admits for eight years, and marginally higher scores for the remaining four years. A t-test shows that there is no statistically significant difference between the average ACT scores of priority-designated and non-priority admits at the 5 percent level.⁵⁷⁷

338. The results combined in Appendix 4 Tables 7-9 establish that, on average, from 2011-2023, MIT's priority-designated applicants (a) were admitted at higher rates than non-priority-designated applicants, and (b) were admitted with similar average standardized test scores than non-priority-designated students. I conclude from these data that some factor other than average ACT or SAT scores was positively influencing the admission of priority-designated students.

D. Notre Dame

339. In my review of Notre Dame's structured data, I observed designations assigned to applicants called (among others) "University Interest," "University Relations A, B, B-, and C," and "President" (collectively "priority designations").⁵⁷⁸

575. $47 / 1600 = 0.0293$.

576. A t-test of all admits' SAT scores by priority designation shows that for each year with available data, the SAT scores are not significantly different at the 5 percent significance level. *See* my workpapers for details.

577. A t-test of all admits' ACT scores by priority designation shows that for each year with available data, the ACT scores are not significantly different at the 5 percent significance level. *See* my workpapers for details.

578. *See* ND_STRUCTURED_000014.

340. Notre Dame provided admissions data for the following categories from 2005 through 2022.⁵⁷⁹ Appendix 4 Table 10 compares admissions rates by academic year for “University Interest,” “University Relations,” “President,” and Non-Priority applicants.⁵⁸⁰

APPENDIX 4 TABLE 10: NOTRE DAME ADMISSIONS RATE BY PRIORITY DESIGNATION

Academic Year	University Interest	Any University Relation	President	Non-Priority Applicants
2005	51.0%	60.6%	46.2%	30.6%
2006	35.2%	52.0%	72.7%	26.4%
2007	43.6%	58.0%	43.9%	23.2%
2008	43.9%	57.1%	44.8%	25.6%
2009	47.2%	57.3%	47.6%	27.6%
2010	44.4%	58.4%	40.0%	27.6%
2011	42.5%	51.4%	37.0%	23.4%
2012	52.4%	54.0%	27.3%	22.4%
2013	-	54.3%	46.7%	21.4%
2014	31.3%	46.4%	40.0%	20.4%
2015	54.5%	49.7%	31.4%	19.0%
2016	-	61.4%	32.4%	18.3%
2017	36.4%	64.0%	41.8%	18.3%
2018	38.1%	-	33.3%	17.6%
2019	-	-	46.9%	15.6%
2020	-	-	53.1%	18.7%
2021	-	-	22.9%	14.8%
2022	-	-	41.5%	12.7%

Notes: The “Non-Priority Applicants” category includes all non-priority-designated applicants. I have excluded admission rates in years where there were less than 10 students in a priority-designation category. The “Any University Relation” category includes any student with a “University Relations ‘A’ Rating,” “University Relations ‘B’ Minus Rating,” “University Relations ‘B’ Rating,” or “University Relations ‘C’ Rating.”

341. After examining these findings, I conducted two t-tests to compare the admission rate of priority-designated applicants with the admission rate of all applicants. I found that applicants between 2005 and 2017, with any University Relations designation, were admitted at a statistically significant higher rate than non-priority-designated applicants in every single year.⁵⁸¹ Further, I found

579. *Id.*

580. ND_STRUCTURED_000014; ND_STRUCTURED_000012. The academic year refers to the year the applicant was admitted. The academic year 2006-2007 is referred to as academic year 2006.

581. Statistical significance was tested at the $p = 0.05$ level. See my workpapers for details.

that applicants with a “President” priority designation were admitted at a higher rate than non-priority applicants in every year and at a statistically significant higher rate for 12 out of 17 years between 2005 and 2022.⁵⁸² Where the results are not statistically significant, the “President” priority applicants are still accepted at a higher rate than non-priority applicants.

342. Based on these findings, I conclude that a statistically significant difference exists between the admission rates of priority-designated applicants (“University Interest,” “University Relations,” and “President”) when compared to the admission rates of non-priority-designated applicants.⁵⁸³ Notre Dame admits priority-designated applicants at a substantially higher rate than the corresponding admissions rate of its non-priority-designated applicants.

343. I also test whether the priority-designated applicants who were admitted had higher standardized test scores relative to the non-priority-designated applicants who were admitted. To do this, I use Notre Dame’s test score data, which provides SAT and ACT scores for Notre Dame applicants. I compare the average SAT and ACT scores for admits with “University Relation” designations and admits without any priority designation by academic year. Appendix 4 Table 11 and Appendix 4 Table 12 provide the average SAT and ACT scores respectively for these two groups of Notre Dame admits from 2007-2016.⁵⁸⁴

582. See my workpapers for details.

583. I found a statistically significant (at the $p = 0.05$ level) difference between the admission rates of students with any “University Relations” designation and students with no priority designation for every year between 2005 and 2017.

584. Due to the low sample size in the “President” and “University Interest” categories, I have compared only non-priority-designated admits and admits with any “University Relation” designation.

APPENDIX 4 TABLE 11: COMPARING SAT SCORES FOR PRIORITY-DESIGNATED AND NON-PRIORITY-DESIGNATED NOTRE DAME ADMITS (2007-2016)

Year	No Priority Designation	Any University Relation
2007	1,952	1,870
2008	2,023	1,911
2009	2,054	1,939
2010	2,081	1,994
2011	2,089	2,038
2012	2,100	2,037
2013	2,112	2,039
2014	2,117	2,067
2015	2,116	2,053
2016	2,110	2,083

Notes: The “Any University Relation” category includes any student with a “University Relations ‘A’ Rating,” “University Relations ‘B’ Minus Rating,” “University Relations ‘B’ Rating,” or “University Relations ‘C’ Rating.”

APPENDIX 4 TABLE 12: COMPARING ACT SCORES FOR PRIORITY-DESIGNATED AND NON-PRIORITY-DESIGNATED NOTRE DAME ADMITS (2007-2016)

Year	No Priority Designation	Any University Relation
2007	31.5	30.2
2008	31.8	30.4
2009	32.0	30.7
2010	32.2	31.0
2011	32.3	31.6
2012	32.3	31.5
2013	32.6	31.8
2014	32.9	31.9
2015	32.8	31.9
2016	32.9	31.8

Notes: The “Any University Relation” category includes any student with a “University Relations ‘A’ Rating,” “University Relations ‘B’ Minus Rating,” “University Relations ‘B’ Rating,” or “University Relations ‘C’ Rating.”

344. The average test scores in Appendix 4 Table 11 and Appendix 4 Table 12 demonstrate that in every year between 2007 and 2016, the average SAT and ACT scores were higher for the non-priority-designated admitted cohort. The results from Appendix 4 Tables 10-12 demonstrate that

priority-designated applicants had a higher admission rate than non-priority-designated applicants between 2007 and 2016, while on average the admitted priority-designated applicants had lower standardized test scores. I also found that the mean ACT scores for non-priority-designated applicants were statistically significantly higher than for priority-designated applicants for every year between 2007 and 2016.⁵⁸⁵ The mean SAT scores for non-priority-designated applicants were statistically significantly higher than for priority-designated applicants between 2007 and 2015.⁵⁸⁶ This indicates that priority-designated students had lower or not statistically different test scores for each year of available data when compared to non-priority-designated applicants.

345. Given the results of my analyses in Appendix 4 Tables 10-12, the Notre Dame admissions data show in a statistically significant way that (a) Notre Dame priority-designated applicants had significantly higher admission rates than non-priority-designated applicants and (b) Notre Dame priority-designated applicants were admitted with lower scores on average compared to non-priority-designated applicants. I conclude from these data that some factor other than average ACT or SAT scores is positively influencing the admission of priority-designated applicants.

E. Penn

346. In my review of Penn's structured data, I observed designations assigned to applicants called (among others) "SI," "special interest," "Bona Fide Special Interest," "bonafide_si," "BSI," and "BSI-A" (collectively "special-interest designations").

347. Based on my review of the structured data provided by Penn, I have observed the following regarding the special-interest designations contained therein:

348. I could not find special-interest designations among Penn's structured data for the years 2016 through 2018.

585. Statistical significance found at the $p = 0.05$ level. *See* my workpapers for details.

586. Statistical significance found at the $p = 0.05$ level. *See* my workpapers for details.

349. For admissions years 2003 to 2009:

- a. Penn admitted approximately 18.42 percent of undergraduate applicants. I have also verified my results by comparing them to public data and published sources.⁵⁸⁷
- b. Penn admitted approximately 65.69 percent of applicants with a special-interest designation.
- c. Penn admitted approximately 17.62 percent of applicants without a special-interest designation.

350. For admissions years 2010 to 2015:

- a. Penn's admission rate for all applicants was 11.72 percent.
- b. Penn's admission rate for applicants with a BSI-A designation was 86.88 percent.
- c. Penn's admission rate for applicants without any special-interest designation was 11.20 percent.

351. For admission year 2019:

- a. Penn's admission rate for all applicants was 7.56 percent.
- b. Penn's admission rate for applicants with a BSI-A designation was 73.18 percent.
- c. Penn's admission rate for applicants without any special-interest designation was 7.14 percent.

352. The table below reflects the data I referenced above, where "Any SI Designation" from 2003 to 2009 includes "bonafide SI," "Dean Admission SI," and "Potential SI" students.

587. See, e.g., Sara Levine, *Admission Rates Drop at Penn, Ivies*, DAILY PENNSYLVANIAN (Apr. 11, 2003), https://www.thedp.com/article/2003/04/admission_rates_drop_at_penn_ivies. My results are consistent with this drop over time. I have verified the Penn admissions rates as reported in the Department of Education's Integrated Postsecondary Education Data System (IPEDS). For example, I have calculated a 7.6 percent admissions rate in 2019 from the Penn structured data. My analysis of the IPEDS data shows a corresponding 2019 Penn admissions rate of 7.7 percent, a nearly identical figure to that obtained from the structured data. My calculation also corresponds closely with the 7.4 percent acceptance rate reported by the Daily Pennsylvanian in March 2019. See, Gillian Diebold, *Penn admits a record-low 7.44 percent of applicants to the Class of 2023*, DAILY PENNSYLVANIAN (Mar. 28, 2019), <https://www.thedp.com/article/2019/03/penn-acceptance-ivy-league-regular-decision-admissions-class-2023>. As the paper explains, "The Daily Pennsylvanian is the University of Pennsylvania's independent student media organization." See also *About The Daily Pennsylvanian, Inc.*, DAILY PENNSYLVANIAN, <https://www.thedp.com/page/about> (last visited Oct. 2024).

APPENDIX 4 TABLE 13: COMPARING ADMISSIONS RATES FOR SPECIAL INTEREST APPLICANTS
COMPARED TO NON-SPECIAL-INTEREST APPLICANTS

Category	2003-09 Admission Rates	2010-15 Admission Rates	2019 Admission Rates
Non-SI Designation	17.62%	11.20%	7.14%
Any SI Designation	65.69%	---	---
BSI-A Designation	N/A	86.88%	73.18%

Notes: For the 2010-2015 admission rates and 2019 admission rates, the Non-SI Designation includes all applicants without a BSI designation.

353. Based on these findings, I conclude that a statistically significant difference exists between the admission rates of students with the “BSI-A” special interest tag or any SI designation when compared to the admission rates of non-special-interest students.⁵⁸⁸ Penn admitted students with a special-interest designation at a substantially higher rate than the corresponding admission rate of both its entire undergraduate applicant pool and students without a special-interest designation.

354. I also test whether the special interest designated applicants who were admitted had higher standardized test scores relative to the non-special-interest applicants who were admitted. To do this, I use Penn’s test score data, which provides SAT and ACT scores for Penn applicants. For each admitted group, I calculate the average SAT and ACT score per academic year. Appendix 4 Table 14 and Appendix 4 Table 15 provide the average SAT and ACT scores respectively for Penn admits from 2003-2009.

588. For every year between 2003 and 2015, I find a statistically significant difference in the admission rates between special-interest and non-special-interest students. Between 2003 and 2009 I test the admission rates of “Any SI” students compared to students without any SI designation. Between 2010 and 2015 I test the admission rates of “BSI-A” students compared to students without any BSI designation.

APPENDIX 4 TABLE 14: COMPARING SAT SCORES FOR SI AND NON-SI PENN ADMITS
(2003-2009)

Academic Year	Special Interest Admits	All Other Admits
2003	1,380	1,422
2004	1,391	1,427
2005	1,384	1,430
2006	2,084	2,128
2007	2,082	2,137
2008	2,118	2,150
2009	2,129	2,166

Notes: From 2003 to 2005, only students with a 1600 scale SAT score are considered. From 2006 to 2009, only students with a 2400 scale SAT score are considered.

APPENDIX 4 TABLE 15: COMPARING ACT SCORES FOR SI AND NON-SI PENN ADMITS
(2003-2009)

Academic Year	Special Interest Admits	All Other Admits
2003	-	-
2004	-	-
2005	-	-
2006	29.9	30.9
2007	30.6	31.1
2008	31.0	31.5
2009	31.4	32.0

Notes: No ACT scores are available in Penn's admissions data from 2003 to 2005.

355. The average test scores in Appendix 4 Table 14 and Appendix 4 Table 15 demonstrate that in every year between 2003 and 2009, the average SAT and ACT scores were higher for the non-special-interest admitted cohort.⁵⁸⁹ Several of these results are statistically significant, but for those that are not, the results indicate that the special-interest students did not score better than the non-special-interest students. The results from Appendix 4 Tables 13-15 demonstrate that special interest

589. The mean SAT score for non-special-interest Penn admits was statistically significantly ($p = 0.05$) higher during all the years between 2003 and 2009. The mean ACT score for non-special-interest Penn admits was statistically significantly ($p = 0.05$) higher for the years 2006 and 2009. *See* my workpapers for details.

applicants had a significantly higher admission rate than non-special-interest applicants between 2003 and 2009, while on average the admitted special interest students had lower test scores.

356. I replicate the previous analysis for the years 2010-2015 and the year 2019 in Appendix 4 Table 16 and Appendix 4 Table 17 below.

APPENDIX 4 TABLE 16: COMPARING SAT SCORES FOR BSI-A AND NON-SI PENN ADMITS
(2010-2015, 2019)

Academic Year	BSI-A Admits	Non- BSI Admits
2010	2,135	2,170
2011	2,124	2,172
2012	2,144	2,178
2013	2,121	2,188
2014	2,102	2,180
2015	2,093	2,182
2019	1,491	1,497

Notes: From 2010 to 2015, only students with a 2400 scale SAT score are considered. In 2019, only students with a 1600 scale SAT score are considered. Special-interest designations are unavailable from 2016 to 2018.

APPENDIX 4 TABLE 17: COMPARING ACT SCORES FOR BSI-A AND NON-SI PENN ADMITS
(2010-2015, 2019)

Academic Year	BSI-A Admits	Non- BSI Admits
2010	30.8	31.8
2011	30.1	31.8
2012	31.1	32.0
2013	30.4	32.3
2014	32.2	32.2
2015	31.0	32.5
2019	33.7	33.6

Notes: Special-interest designations are unavailable from 2016 to 2018.

357. The results from Appendix 4 Table 16 and Appendix 4 Table 17 replicate the earlier finding for 2003-2009. For the years between 2010 and 2015, the non-special-interest Penn admits have either an equal or higher average SAT and ACT score compared to “BSI-A” special interest

admits.⁵⁹⁰ Only the ACT scores in 2019 show “BSI-A” applicants having higher average scores (without statistical significance).⁵⁹¹ At the same time, as shown in Appendix 4 Table 13, the “BSI-A” cohort had a 74 percentage-point higher admission rate from 2010 to 2015 compared to non-special-interest applicants.

358. Given the results of my analyses in Appendix 4 Tables 13-17, the evidence from the Penn admissions data indicates that (a) Penn special-interest designated applicants had a statistically significantly higher admission rate compared to non-special-interest applicants, and (b) Penn special-interest applicants were admitted with statistically significant lower or equal scores to non-special-interest applicants. I conclude from these data that some factor other than average ACT or SAT scores is positively influencing the admission of special-interest applicants.

590. Between 2010 and 2015, the mean SAT score for non-special-interest Penn admits was statistically significantly ($p = 0.05$) higher than BSI-A admits during the years 2014 and 2015. The mean ACT score for non-special-interest Penn admits was statistically significantly ($p = 0.05$) higher than BSI-A admits during the years 2011, 2013, and 2015. *See* my workpapers for details.

591. This difference in mean ACT scores is not significant at the $p = 0.05$ level. *See* my workpapers for details.